

# Rational Decision Support with a Natural Language Dialogue System

Vom Fachbereich Informatik der  
Technischen Universität Darmstadt  
genehmigte

## **Dissertation**

zur Erlangung des akademischen Grades Dr. rer. nat.

vorgelegt von

**Dipl.-Inform. Daniel O. Mäurer**  
geboren in Frankfurt am Main

Tag der Einreichung: 7. September 2015

Tag der Disputation: 27. Oktober 2015

Referenten: Prof. Dr. Karsten Weihe, TU Darmstadt  
Prof. Dr. Johannes Fürnkranz, TU Darmstadt

Darmstadt 2016  
D17



# Abstract

---

In the past decades, technological advance has led to a revival of natural language dialogue systems (or conversational agents). Now, conversational agents are used for professional purposes such as customer support, marketing, e-learning, and tutoring purposes, just to mention a few. They also have found their way into everyday life, e.g. as speech-based personal assistants on mobile devices or as freely available text-based information retrieval systems in the web. However, there are still many unexplored application areas for conversational agents. Personal coaching, for example, is a promising field. Coaching has been successfully applied across many areas, for example in human resource development, team-building, and improvement of various individual skills such as leadership, communication or sales, and to support humans with rational decision-making. Nevertheless, personal coaching by a human coach or counsellor is very expensive and therefore only available to a limited circle of individuals in organizations. So far, conversational agents have not been extensively used for personal coaching purposes.

This work intends to explore the potential of conversational agents for simulated human coaching or consulting. Therefore, we have developed VPINO, a text-based dialogue system, intended for holding structured coaching conversations in German language. This work provides a detailed description of VPINO. A general maxim of many modern coaching approaches is to regard the client as the expert for the relevant problem, rather than seeing the coach as the expert for the client's problem. VPINO aims to open up a new perspective and help the "expert" develop his/her own solution. Instead of giving advice or pushing the client into any certain direction, VPINO uses the technique of Socratic questioning, i.e. targeted questions to support clients to reflect on their goal accomplishment process. VPINO is an expert on structuring the user's implicit knowledge and merely assists the user like a human coach. The intelligence or value of our dialogue system is not in its ability in detailed understanding what the user says and will therefore not provide answers to particular questions. The key is that, nevertheless, VPINO understands yet enough to keep control over the conversation, in a way that will not be noticed by the user. To do so, the system determines the dialogue act of each user utterance, i.e. the function of that utterance for the conversation. A suitable response is selected from a set of preformulated utterances, based on a model of potential dialogue act sequences. VPINO has been

developed for application in the particular scenarios of (1) training transfer coaching for communication skills and (2) rational decision support. The effectiveness and usefulness of VPINO in both scenarios are evaluated in a series of user studies.

For the scenario of *training transfer coaching*, we conducted a field experimental study that evaluates the effectiveness of coaching, right after participants received an online communication training. Training transfer, the application of newly acquired skills in everyday work, requires a maximum of support and reflection. Our results suggest that a computer-based coaching with VPINO can effectively provide this support and improve the participants communication skills. Moreover, we explored the influence of the user's personality and behaviour when talking to the computer based coach. Whereas users that are more "open to new experience" particularly benefit from chat coaching with VPINO, more sceptical and less cooperative users do not profit from using VPINO.

For the scenario of *rational decision coaching*, VPINO was evaluated in a series of user studies. VPINO guides the users through a process of structured decision-making on a distinctly rational basis, in the spirit of Benjamin Franklin's method "Pros and Cons". Participants were free to talk about a decision problem of their own choice. Whereas VPINO could successfully help participants with rational problems, the conversation was less effective with more emotionally stressful problems. However, participants with a generally less structured decision making approach particularly benefit from using VPINO. Participants that are generally more open to receiving support from an artificial coach evaluated the usefulness of the conversation higher than sceptical users. However, a high level of motivation and cooperativeness is required for fruitful conversation. In general, user expectations, behaviour, and willingness to cooperate have an influence on how users interact with the virtual coach. The use of chat language greatly varies between users in many dimensions, e.g. grammatical and syntactical correctness, style, length of answers, and politeness.

This thesis shows that a natural language dialogue system can serve as a highly available, cheap, low threshold solution to support a large target group, where a human personal coach is not available.

# Zusammenfassung

---

Computerbasierte Dialogsysteme in natürlicher Sprache haben Dank des technischen Fortschritts der letzten Jahrzehnte, z.B. im Bereich mobiler Endgeräte, dem Internet, und im Natural Language Processing, eine Renaissance erfahren. Mittlerweile sind sie neben professionellen Einsatzgebieten auch im Alltag vieler Menschen angekommen, z.B. als sprach- oder chatbasierte Assistenzsysteme auf mobilen Endgeräten. Dennoch gibt es viele bislang unerschlossene Anwendungsgebiete für Dialogsysteme. Ein Beispiel dafür ist professionelles Coaching. Coaching wird in vielen Bereichen wie der Personalentwicklung, Teambuilding, zur Verbesserung individueller Fähigkeiten, sowie zur rationalen Entscheidungsfindung angewandt. Ein professioneller Coach oder Berater ist kostspielig und steht daher oft nur einem begrenzten Kreis von Personen in Organisationen zur Verfügung. Bisher werden Chatbots oder vergleichbare Dialogsysteme noch nicht umfassend für Coachingzwecke eingesetzt.

Mit der vorliegenden Arbeit soll das Potenzial von Dialogsystemen für simulierte menschliche Coachings erkundet werden. In diesem Zusammenhang haben wir VPINO entwickelt. VPINO ist ein chat-basiertes Dialogsystem für strukturierte Coaching-Gespräche in deutscher Sprache, dessen Funktionsweise in dieser Arbeit detailliert beschrieben wird. Ein Grundsatz vieler moderner Coaching-Ansätze ist es, den Klienten als den eigentlichen Experten für sein jeweiliges Problem zu betrachten. Der Coach dagegen übernimmt die Rolle des objektiven Beraters, ohne dabei den Klienten durch konkrete Ratschläge in eine bestimmte Richtung zu beeinflussen. VPINO handelt genau in diesem Sinne. VPINO ist ein Experte für die Strukturierung des beim Klienten implizit vorhandenen Wissens und unterstützt diesen genauso wie ein menschlicher Coach. Dabei nutzt der Coach die Technik des „Sokratischen Fragens“, d.h. den Klienten durch gezieltes Fragen bei der Erreichung seiner Ziele zu unterstützen. Der Mehrwert des Dialogsystems liegt dabei nicht in seiner Fähigkeit, ein detailliertes inhaltliches Verständnis des Problems zu entwickeln. VPINO gibt keine Antworten auf inhaltliche Fragen des Nutzers. Grundprinzip dabei ist, dass VPINO dennoch genug versteht, um die Kontrolle über den Gesprächsverlauf zu behalten. Dafür bestimmt VPINO die Dialogakte der Nutzeräußerungen, um eine passende Reaktion auszuwählen. VPINO steuert das Gespräch möglichst in einer Art und Weise, dass der Nutzer dieses nicht bemerkt.

VPINO wurde für die Anwendung im Bereich (1) *Training Transfer-Coaching* für Kommunikationsfähigkeiten als auch für (2) *rationales Entscheidungs-Coaching* optimiert. Die Wirksamkeit und Nützlichkeit des Systems wurde in einer Reihe von Benutzerstudien ausgewertet, welche in dieser Arbeit vorgestellt werden.

Für das Szenario *Training-Transfer-Coaching* wurde eine experimentelle Feldstudie durchgeführt, um die Wirksamkeit von computergesteuerten Transfer-Coachings im direkten Anschluss an ein Online-Kommunikationstraining auszuwerten. Der Transfer von neu erworbenen Fähigkeiten in den (Arbeits-)Alltag erfordert ein Höchstmaß an Unterstützung und Reflexion. Die Ergebnisse der Studie legen nahe, dass ein computergesteuertes Coaching mit VPINO die Nutzer effektiv unterstützen und zur Verbesserung ihrer Kommunikationsfähigkeiten beitragen kann. Darüber hinaus wurde der Einfluss der Persönlichkeit sowie des Nutzerverhaltens auf den Erfolg des Coachings erforscht. Diejenigen Nutzer, die allgemein mehr „Offenheit für neue Erfahrungen“ zeigen, profitieren besonders von Coachings durch VPINO. Skeptische und wenig kooperative Nutzer dagegen profitieren im Vergleich weniger.

Für das zweite Szenario, dem *rationalen Entscheidungs-Coaching*, führt VPINO den Anwender durch einen Prozess der strukturierten Entscheidungsfindung auf rationaler Basis, im Geiste von Benjamin Franklins Methode „Pros und Cons“. In der vorliegenden Arbeit werden dazu zwei durchgeführte Nutzerstudien vorgestellt. Die Teilnehmer der Studien sprachen mit dem Entscheidungscoach über ein Entscheidungsproblem ihrer Wahl. Es wurde untersucht, wie die grundsätzliche Einstellung der Nutzer sowie deren übliche Vorgehensweise bei der Entscheidungsfindung Einfluss auf den Erfolg des Coachinggesprächs mit VPINO haben. Während VPINO den Teilnehmern mit eher rationalen Problemen weiterhelfen konnte, waren Gespräche mit einem Fokus auf persönlichere, emotional belastende Probleme dagegen weniger erfolgreich. Insbesondere profitierten Teilnehmer mit einem weniger strukturierten Ansatz bei der Entscheidungsfindung. Teilnehmer, die offener für computergestützte Coaches sind, bewerteten die Nützlichkeit des Gesprächs höher als skeptische Nutzer. Selbstverständlich sind ein gewisses Maß an Motivation und Mitarbeit die Grundvoraussetzung fruchtbarer Gespräche. Im Allgemeinen haben Nutzererwartungen, Nutzerverhalten, sowie Bereitschaft zur Zusammenarbeit einen großen Einfluss auf die Interaktion der Benutzer mit dem virtuellen Coach. Die verwendete Chat-Sprache variiert stark zwischen einzelnen Nutzern, z.B. grammatikalische und syntaktische Korrektheit, Stil, Länge und Ausführlichkeit der Antworten sowie Umgangston und Respekt. Dadurch wird die Interpretation von Chat-Sprache im Gegensatz zu sonstiger geschriebener Sprache deutlich erschwert.

Mit der vorgelegten Arbeit konnte gezeigt werden, dass computergestütztes Coaching eine preiswerte und schnell verfügbare Alternative für Situationen darstellen könnte, in denen ein menschlicher Coach nicht zur Verfügung steht.

# Acknowledgments

---

First, I would like to express my gratitude to my advisor Karsten Weihe for giving me the chance to write this thesis and supporting me at all times throughout my work.

Moreover, I would like to thank Anna Bruns for her support with the evaluation of the user studies and co-authoring the publication on training transfer coaching.

I also want to thank Felix Sehr and Wibke Ellen for bringing in their ideas and their help with authoring and testing the developed software. Thanks to Felix Gündling for his help with software development, and Moses Githinji for recruiting participants for my user studies. Furthermore, I would also like to thank my colleagues at the Algorithmics group.

A special thanks goes to my Family, especially my parents and my brother, who helped and inspired me my whole life. Finally, I would like to express my deepest gratitude to Tatjana, for always being there and cheering me up when needed.





# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Structure of this Work . . . . .	2
1.3	What is Coaching? . . . . .	3
1.3.1	Definition of Coaching . . . . .	3
1.3.2	A Goal Oriented Process . . . . .	4
1.3.3	Virtual Coaching and E-Coaching . . . . .	4
1.4	A Conversational Agent as a Coach . . . . .	5
1.4.1	VPINO . . . . .	6
1.4.2	Socratic Questioning . . . . .	6
1.4.3	Proactive Behaviour . . . . .	7
1.4.4	Natural Language Understanding . . . . .	8
1.5	Background and Related Work . . . . .	8
1.5.1	Artificial Intelligence . . . . .	8
1.5.2	Chatbots . . . . .	9
1.5.3	AIML and ALICE . . . . .	10
1.5.4	Façade . . . . .	11
1.5.5	Conversational Agents in Education . . . . .	12
1.5.6	Conversational Agents for Coaching . . . . .	13
1.5.7	Information Retrieval / Question Answering Systems . . . . .	13
1.6	Contributions and Findings . . . . .	14
1.6.1	Proof of Concept . . . . .	14
1.6.2	Scenario I: Training Transfer . . . . .	14
1.6.3	Scenario II: Rational Decision-Making . . . . .	15

## CONTENTS

<b>2</b>	<b>Background and Concepts Behind VPINO</b>	<b>17</b>
2.1	Big Picture of Vpino's Background . . . . .	18
2.2	Dialogue Act Theory . . . . .	19
2.2.1	Background on Dialogue Acts . . . . .	19
2.2.2	Taxonomies for Chat Conversation . . . . .	22
2.2.3	Dialogue Acts Taxonomy for Structured Coaching . . . . .	22
2.3	Dialogue Act Classification with Vpino . . . . .	24
2.3.1	VPINO Classification Pipeline . . . . .	25
2.3.2	Preprocessing Components . . . . .	25
2.3.3	Rule-Based Text Annotation with Ruta . . . . .	26
2.3.4	The Ruta Rule Language . . . . .	27
2.3.5	Example Dialogue Act Classification Rules . . . . .	28
2.4	Grounding Background . . . . .	30
2.4.1	Grounding Theory . . . . .	30
2.4.2	Methods for Grounding . . . . .	31
2.5	Sub-Dialogues . . . . .	32
2.5.1	Sub-dialogue Organisation . . . . .	33
2.5.2	Grounding/Presentation Phase . . . . .	33
2.5.3	Dialogue Acts Sequence Phase . . . . .	34
2.5.4	Evaluation Phase . . . . .	36
2.5.5	Problem Handling . . . . .	37
2.6	Communication and User Interface . . . . .	37
2.6.1	Timing and Turn-Taking . . . . .	37
2.6.2	Mixed-Initiative Communication . . . . .	39
2.6.3	VPINO Turn-Taking Protocol . . . . .	39
2.6.4	Waiting Time/Presentation Delay . . . . .	41
2.6.5	Visual Delay Hint . . . . .	41
<b>3</b>	<b>Systemic Coaching: Pre-study</b>	<b>43</b>
3.1	Systemic Coaching . . . . .	44
3.2	Vpino Dialogue Plan for Systemic Coaching . . . . .	44
3.3	User Study . . . . .	45
3.3.1	Results . . . . .	45
3.3.2	Naturalness of Conversations . . . . .	45
3.4	Discussion and Conclusions . . . . .	46

<b>4</b>	<b>Training Transfer Coaching: User Study</b>	<b>47</b>
4.1	Background and Motivation . . . . .	48
4.2	Does Personality Affect the Success of Coaching? . . . . .	48
4.3	Simulation-based Training on Real-life Incidents . . . . .	49
4.4	Solution Based Brief Coaching . . . . .	51
4.5	VPINO Dialogue Plan for Brief Coaching . . . . .	51
4.5.1	Initial Conversation . . . . .	51
4.5.2	Recurring Daily Conversations . . . . .	52
4.6	User Study . . . . .	52
4.6.1	Study Design . . . . .	52
4.6.2	Measures . . . . .	54
4.6.3	Control Variables . . . . .	55
4.7	Results . . . . .	56
4.7.1	Descriptives . . . . .	56
4.7.2	Quantitative Results . . . . .	56
4.7.3	Qualitative Results . . . . .	58
4.8	Discussion and Conclusions . . . . .	59
<b>5</b>	<b>Rational Decision Coaching with VPINO</b>	<b>61</b>
5.1	Background and Motivation . . . . .	62
5.1.1	Rational Decision Making . . . . .	62
5.1.2	Background on Decision Support . . . . .	62
5.2	Benjamin Franklin's Method . . . . .	63
5.2.1	Pros and Cons . . . . .	63
5.2.2	Moral Algebra . . . . .	63
5.2.3	Discussion of Franklin's Approach for Coaching . . . . .	64
5.3	Human Decision-Making Patterns . . . . .	65
5.3.1	Vigilance and Hypervigilance . . . . .	66
5.3.2	Heuristic and Intuitive Decision-Making . . . . .	66
5.4	Rational Decision Dialogue Structure . . . . .	67
5.4.1	Structured Knowledge Memory for Decision Coaching . . . . .	67
5.4.2	Attributes . . . . .	68
5.4.3	Dialogue Plan . . . . .	69
5.4.4	Stage I - Problem Framing and Goal Setting . . . . .	69
5.4.5	Stage II - Options and Arguments Collection . . . . .	70

## CONTENTS

5.4.6	Stage III - Review and Next Steps . . . . .	70
5.5	Pairwise Option Comparison . . . . .	73
5.5.1	Option Characteristics . . . . .	73
5.5.2	Option pair selection . . . . .	75
5.5.3	Paraphrasing Option Pairs . . . . .	75
5.6	Semantic Reference Resolution on Options . . . . .	76
5.6.1	Baseline approach . . . . .	77
5.6.2	Normalization . . . . .	77
5.6.3	Semantic Similarity Measures . . . . .	77
5.6.4	Evaluation . . . . .	78
<b>6</b>	<b>Rational Decision Coaching: User Study</b>	<b>79</b>
6.1	Setting . . . . .	80
6.1.1	Study Design . . . . .	80
6.1.2	Measures . . . . .	80
6.2	Results . . . . .	82
6.2.1	Openness and Attitude . . . . .	83
6.2.2	Hypervigilance . . . . .	83
6.3	Qualitative Evaluation . . . . .	84
6.3.1	Reference Resolution . . . . .	85
6.3.2	Dialogue Act Classification . . . . .	86
6.3.3	Expectations and Disappointment . . . . .	86
6.4	Discussion and Conclusions . . . . .	87
<b>7</b>	<b>Rational Decision Coaching: Follow-up Study</b>	<b>89</b>
7.1	Motivation and Improvements on VPINO . . . . .	90
7.2	Setting . . . . .	90
7.2.1	Study Design . . . . .	90
7.2.2	Measures . . . . .	91
7.3	Results . . . . .	92
7.3.1	Motivation, Cooperation and Attitude . . . . .	93
7.3.2	Hypervigilance . . . . .	93
7.3.3	Personal vs. Professional Problems . . . . .	93
7.3.4	Age and Gender . . . . .	96
7.4	User Feedback . . . . .	96
7.4.1	Positive Feedback . . . . .	96

7.4.2	Negative Feedback . . . . .	97
7.5	Discussion and Conclusions . . . . .	98
<b>8</b>	<b>Software Architecture</b>	<b>99</b>
8.1	Communication and Software Architecture of VPINO . . . . .	99
8.1.1	User Interface . . . . .	99
8.1.2	Publish/Subscribe Messaging Server . . . . .	101
8.1.3	VPINO Session-Manager . . . . .	101
8.1.4	Dialogue Act Classifier . . . . .	101
8.1.5	Speech-based Communication . . . . .	101
8.2	Authoring DEV Tool . . . . .	102
8.2.1	Sub-Dialogue Editor . . . . .	104
8.2.2	Dialogue-Plan Editor . . . . .	104
<b>9</b>	<b>Discussion and Conclusions</b>	<b>107</b>
	<b>List of Figures</b>	<b>122</b>
	<b>List of Tables</b>	<b>123</b>
<b>A</b>	<b>Prestudy</b>	<b>125</b>
A.1	Systemic Coaching Questions . . . . .	125
<b>B</b>	<b>Training Transfer</b>	<b>127</b>
B.1	Communication Training Tool . . . . .	127
B.2	Conventional Transfer Strategy . . . . .	128
<b>C</b>	<b>Decision Making</b>	<b>131</b>
C.1	Benjamin Franklin's Letter to Joseph Priestly . . . . .	132
C.2	Example Decision Support Conversation with VPINO . . . . .	133
C.3	Teilnahmebedingungen Nutzerstudie . . . . .	137

## CONTENTS

## CHAPTER 1

# Introduction

---

### 1.1 Motivation

Machines that are able to act, think or talk like humans have been inspiring a lot of authors, philosophers and researchers ever since. A lot of literature on the topic of intelligent machines has become famous in the last centuries, from Mary Shelley's "Frankenstein" to Asimov's "I, Robot" or the intelligent computer system HAL in Arthur C. Clarke's "A Space Odyssey". Nevertheless, the appearance of machines with human like skills seems to be behind schedule. Still, we do not see robots roaming our streets. There is still an ongoing discussion whether strong/deep artificial intelligence that matches the full range of human cognitive skills is possible or not.

However, research on specific artificial intelligence related tasks has made a huge progress, for example in the fields of image processing, knowledge engineering and especially the field of computational linguistics and natural language processing. Recent computer systems are now able to process large amounts of unstructured data, such as text, and transform relevant information into structured data. Speech recognition and text-to-speech software enabled communication interfaces that allow humans to communicate with a computer in natural language. Other technological development, such as the appearance of the internet and mobile devices, opened new perspectives for communication systems such as conversational agents. A conversational agent (or dialogue system) is a computer program that uses natural language to engage humans in a conversation, either spoken or written.

There are plenty of text based dialogue systems used in practice for domain specific tasks such as sales, marketing, FAQ<sup>1</sup> or customer support. Also, a few are used for e-learning and tutoring purposes. Still, the majority of these task-oriented systems used

---

<sup>1</sup>Frequently asked questions

in practice apply rule-based approaches [Chiticariu et al., 2013]. Most of the information retrieval systems, in contrast, apply statistical or machine learning methods to process natural language.

Personal assistants and information retrieval systems for mobile devices, such as Apple’s “Siri”, “Google Now” or Microsoft’s “Cortana”, have found their way into everyday life by providing knowledge and information from the web.

Nevertheless, there is still a gap between natural language interfaces for information retrieval and truly intelligent, cognitive agents. Holding and keeping up a coherent conversation, even in written conversation, is still a big challenge. Conversational agents for entertainment, called chat-bots, are still not able to trick humans into believing they are talking to a human at all.

However, there are still many unexplored application areas for conversational agents. Personal coaching, for example, is a promising yet challenging field: The client-centric coaching maxim of “counseling without advice” [Radatz, 2003] involves the guidance of individuals by stimulating questions, rather than by concrete instructions or correct answers. Therefore, a computer based coaching system does not necessarily need to understand the semantic details of the client’s problem domain. We see a chance for a new dialogue system for the purpose of coaching. Such a coaching dialogue system aims to hold a professional, structured conversation instead of entertaining or tricking humans into believing it is human. Thus, we see a perspective for a new dialogue system for the purpose of coaching. So far, conversational agents have not been developed for personal coaching.

## 1.2 Structure of this Work

This thesis is structured as follows. In this first Chapter, the coaching system VPINO is introduced. Furthermore, an introduction into computer based coaching and related work presented. Chapter 2 provides theoretic background and the basic concepts behind of VPINO. Theoretic background includes information about dialogue act and grounding theory. Background is followed by the implementation details of the dialogue act classifier, the sub-dialogue model and communication and turn-taking strategy of VPINO. Chapter 3 presents a pre-study on systemic coaching with an early prototype implementation of VPINO. In Chapter 4, we introduce an implementation of VPINO as a training transfer coach, followed by a user-study on the evaluation of its effectiveness. Chapter 5 gives a detailed description of the implementation of VPINO as a rational decision coach. In Chapter 6 and Chapter 7, two user studies on the effectiveness of VPINO as a rational decision coach are presented. An overview of the software architecture and the authoring tool developed in the course of this thesis is presented in Chapter 8. The thesis is closed with a summary of conclusions and findings, and give an outlook on future work in Chapter 9.



### 1.3 What is Coaching?

In a professional setting, coaching has successfully been applied across many areas such as human resource development, team-building, decision-making or improvement of various individual skills such as leadership, communication or sales, just to name a few. However, there is no coherent definition for the term coaching so far.

Coaching or counselling is often described as a profession rather than a research discipline, although modern approaches originate cognitive and behavioural science (i.e. therapeutic interventions). In a survey on coaching in research and practice, [Grant et al., 2010] report that this is due to the fact that it is mostly performed by non-psychologists the majority of coaches practising today do not use theoretically coherent approaches and scientifically validated techniques and measures.

This section we will give a glance over the common definitions for the term coaching and the resulting process. The term *coaching* is also used interchangeably with counselling, consulting and others. For reasons of consistency, we will keep using the term “coaching” throughout this work. More detailed information about the specific coaching approaches applied in this work, i.e. rational decision coaching and training transfer coaching, will be presented in the corresponding chapters containing the user studies (See Ch. 4 and 5).

#### 1.3.1 Definition of Coaching

The International Coach Federation<sup>2</sup> (ICF), a large professional coaching organization, defines coaching as “ ... partnering with clients in a thought-provoking and creative process that inspires them to maximize their personal and professional potential ...”[ICF, 2015]. According to the ICF, the client (1) should be honoured as the expert in his or her life and work, (2) is believed to be creative and resourceful and (3) has the potential to achieve changes and goals by him/herself. Therefore, the responsibility of a coach in this process is to (1) discover, clarify, and align with what the client wants to achieve, (2) encourage client self-discovery, (3) elicit client-generated solutions and strategies and (4) hold the client responsible and accountable.

[Grant et al., 2010] summarizes the coaching process as a cycle of monitoring, evaluation and modification to create intentional behaviour change.

According to [Hamlin et al., 2008], coaching can be described as a human consultation/intervention intended to support a change process in human organizational behaviour. The role of the coach is to facilitate the clients movement through this self-regulatory cycle by helping the client to develop specific action plans and then to monitor and evaluate progression towards those goals.

---

<sup>2</sup>[www.coachfederation.org](http://www.coachfederation.org)

While many definitions of coaching focus on the roles of the client and the coach, others define coaching as a process for improving problem work performance [Fournies, 1987], or a process of guidance, encouragement, and support to the learner [Redshaw, 2000].

### 1.3.2 A Goal Oriented Process

Whereas the definitions and models emerge from different application areas, it is plausible that they differ. However, all of them agree on a view of coaching as a collaborative relationship formed between coach and client for the purpose of goal attainment. The relationship is characterized as client-centric, meaning the client is valued as the expert that plays an active part in the coaching process.

The goal-oriented, goal-focused or solution-driven paradigm sees the primary function of coaching in fostering the client's self-regulation [Ives, 2008]. According to [Grant, 2003] "Coaching is essentially about helping individuals regulate and direct their interpersonal and intra-personal resources to better attain their goals." [Berg and Szabo, 2005] call this goal-orientated approach also "brief coaching", since it aims to achieve its goals in a comparatively short space of time and normally focusing on a relatively defined issue or goal.

A well established process model for coaching is the GROW model [Whitmore, 2009]. The coaching process involves four consecutive processes: (1) goal setting, (2) examination of real conditions, (3) finding options, alternative strategies or courses of action and (4) implementation of the objectives. In an integrative review of ten coaching models, [Carey et al., 2011] identified four crucial factors close to the GROW Model, (1) coaching relationship, (2) definition of problem/objective, (3) problem solving and (4) the transfer process.

[Ives, 2008] provide a comprehensive overview on the various different coaching paradigms such as humanist, behaviourist, positive psychology, goal-oriented and systemic coaching. A detailed survey of research on the effectiveness and the state of the art in coaching can be found in [Grant et al., 2010].

The work presented in this thesis will focus on the sub-disciplines of rational decision coaching and training transfer coaching employing the paradigm of client-centric goal-oriented coaching.

### 1.3.3 Virtual Coaching and E-Coaching

Research on coaching in the e-learning context, also called *e-coaching*, is still a relatively young discipline. Technological innovations and the increasing demand for highly available coaching that is independent of time or space restrictions were the main reasons for this development. However, coaching was not yet applied in an entirely computer based variant. Although most of the "e-coaching" approaches try to employ modern media for coaching, nearly all of them are limited to technical substitution of face-to-face communication through synchronous or asynchronous text chat, phone coaching or video commu-

nication [Clutterbuck and Hussain, 2010]. Although increasingly supported by technology, these forms of e-coaching always use a human coach in the background.

[Geißler and Metz, 2012] present a “Virtual Goal Attainment Coaching” that consists of telephone-based coaching sessions combined with internet-mediated questions that the client answers with support of the coach.

To our knowledge, there are no natural language systems that focus on decision support or coaching in general. [Schulman et al., 2011] present a virtual counsellor for healthcare behaviour change, using a technique called “motivational interviewing.” Their system is based on a system-directed dialogue planner DTask [Bickmore et al., 2011] for multiple-choice user input. Therefore, it does not support natural language communication.

## 1.4 A Conversational Agent as a Coach

To summarize the coaching definitions from the previous section, goal oriented coaching is a form of collaborative relationship to perform a client-centric process that facilitates goal attainment. The process of goal attainment includes setting of goals, developing action plans and putting these plans into practice. The client-centric view of coaching gives a good perspective for a computer based coach: A coach does neither have to be an expert in the client’s problem domain, nor does he necessarily need to understand the user’s problem or goals at all. The role of the coach is to help the client to structure the problem and guide the process to client-generated solutions and strategies. Accordingly, the client-centric approach achieves goal attainment by utilizing the client’s own problem solving abilities, not by giving advice. The verbalization of the clients thoughts, i.e. writing them down, or explaining them to another person, has been proven to foster reflection, regardless whether the counterpart understands what a person is saying. This positive effect of verbalization, or self-explanation, is also often used as a learning strategy in the educational domain [Ericsson and Simon, 1980, Aleven and Koedinger, 2002]. Therefore, detailed understanding of the underlying problem is not necessarily required. As a consequence, we argue that there is no need of deep artificial intelligence in order to implement an effective coach.

Many of the coaching definitions, such as the GROW model [Whitmore, 2009], describe a somehow formal process for generic problem solving. In this thesis we will show that this generic process can be transformed into a plan for structured conversation. A formal structure may not be suitable for each and every possible coaching scenario, but it can be successfully applied to rational decision coaching or training transfer coaching.

In this section, the idea of adapting the principles of coaching with a conversational agent is discussed. A conversational agent is a computer system that is capable of holding a structured, coherent conversation with a human user in natural language, either spoken, written or embodied (i.e. supported by gestures visual representation).

There are a lot of synonyms or closely related terms for dialogue systems such as chat bots, chatter bots, virtual assistants, chat bot, conversational agents or virtual agents. In this work, we will keep using the terms conversational agent or dialogue system. Also, this work focuses on text based (chat-)communication. In the following, we introduce our dialogue system for coaching and describe the basic conceptual principles. More background information and implementation details will be presented in the following chapter.

### 1.4.1 VPINO

In this work, we introduce VPINO, a text based natural language dialogue system specifically developed for the purpose of holding structured, goal directed coaching conversations. VPINO aims to simulate a human like conversation, including social cues, grounding techniques (See Sect. 2.4) and human turn taking behaviour. It is a mixed initiative system, which means that both conversation partners are allowed to take the initiative at any time. VPINO communicates via a lightweight web interface as displayed in Fig. 1.1.

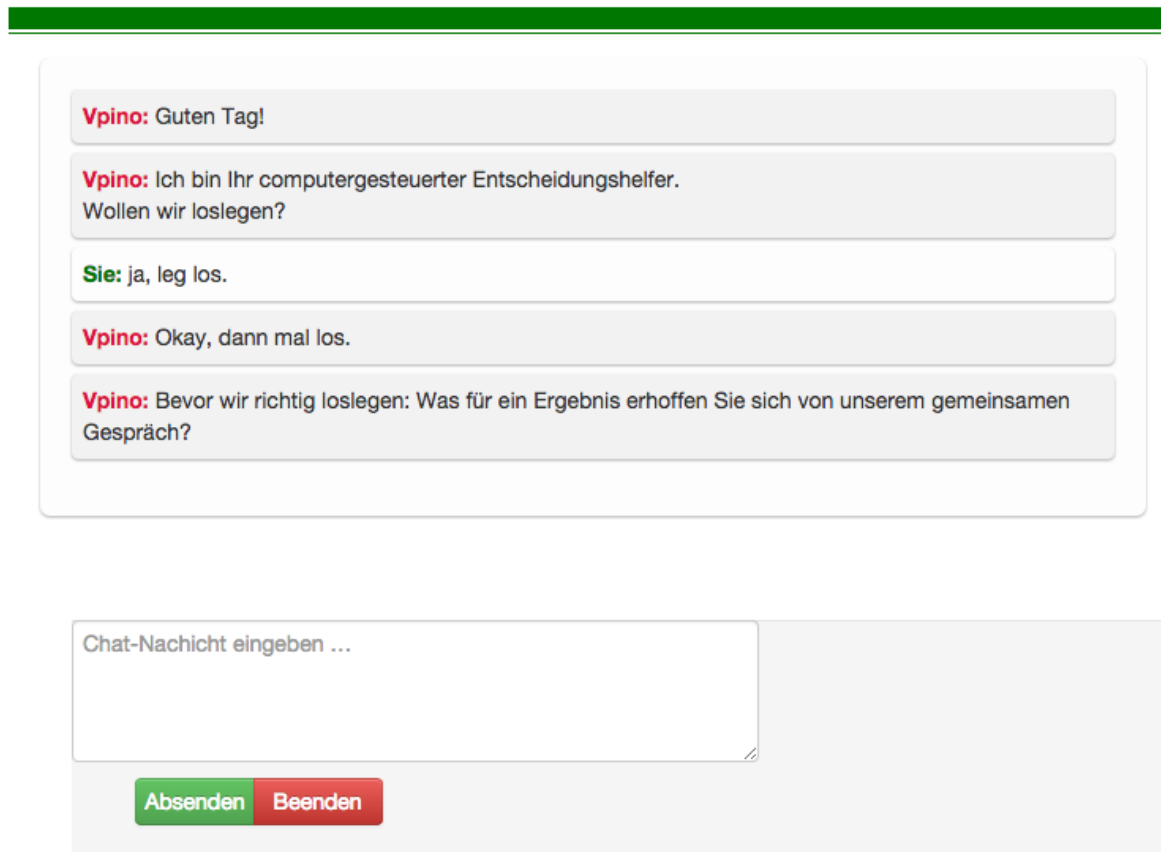
The name VPINO is an artificial term that is based on the Italian word “volpino” that translates to the English adjectives fox-like or vulpine. VPINO thereby refers to cognitive skills often attributed to the animal fox (Latin: “Vulpes”), such as smart, clever, intelligent, cunning or sly.

Following the definitions and basic concepts of coaching discussed in the previous section, VPINO is built on the following principles:

### 1.4.2 Socratic Questioning

As briefly mentioned in the introduction, VPINO applies the method of *Socratic questioning*. Socratic questioning (or Socratic dialoguing), also referred to as *maieutics*, is a general conversation technique that is attributed to the Greek philosopher Socrates. The goal of a Socratic conversation is to explore implicit knowledge, uncover assumptions and follow logical implications by systematic and disciplined questioning. Socratic dialoguing is known to improve critical thinking on the problem subject [Paul and Binker, 1990]. [McGuire, 1960] argues, that the application of the Socratic questioning technique causes people to display more logical consistency.

Although often characterized as a pedagogic method, Socratic questioning is also used in the field of coaching and psychotherapy [Neenan and Palmer, 2001]. In accordance with modern coaching approaches, the client is seen as the expert on the problem subject. Hence, coaching with Socratic questions does not intend to give advice, or push the client into any certain direction, but rather help the “expert” (i.e. the client) to open up a new perspective and to develop his own solution. An ideal Socratic coach would do this by asking targeted questions *only*.



**Figure 1.1:** The web-based user interface with an example beginning of a dialogue. VPINO starts the conversation with a greet and asks if the user is ready to go on. The user agrees, and VPINO moves on to ask for the user's expectations on the coaching conversation.

### 1.4.3 Proactive Behaviour

A major task of VPINO is to guide the client through the coaching process. Therefore, in order to keep control, VPINO needs to be aware of the state and direction of the conversation at all times. To do so, VPINO acts pro-actively: For each specific turn in the conversation, VPINO uses precise preformulated questions and response texts that target a single specific aspect. According to [L'Abbate et al., 2005], pro-actively formulated utterances are a good technique to narrow the range of potential user responses. Thereby, the complexity for the tasks to understand the user responses and to keep track of the conversation state is reduced. As a result, pro-active behaviour enables VPINO to steer the conversation in its desired direction.

#### 1.4.4 Natural Language Understanding

Since a coach does not necessarily need to understand the exact details of a client's utterances, "shallow" natural language understanding capabilities based on dialogue acts are sufficient for directing the course of a coaching conversation. With pro-actively targeted formulations, the communicative function of user utterances, the *dialogue act* an utterance performs, is sufficient for VPINO to continue the conversation properly. Therefore, the general natural language understanding capability is based on dialogue act classification (See Sect. 2.3 for details). Based on the classification of a user utterance, VPINO determines a possible reaction to that specific dialogue act. The user's responses are parsed for semantic information only when needed for continuing a fruitful conversation, depending on the current context and state of the conversation. VPINO is not intended as an information retrieval or question answering system. Therefore, it will never answer user questions or requests related the problem.

### 1.5 Background and Related Work

This section provides a short introduction on text based conversational agents in general and an overview on work related to VPINO. Conversational agents are a rather wide field with a lot of sub-disciplines, such as *question answering* or *information retrieval*, *chatbots*, and others. Holding a structured coaching conversation with a dialogue system differs from many other application scenarios. Thus, in order to avoid misunderstandings, we will point out the differences to these scenarios and existing systems. We distinguish between holding a structured coaching conversation and other application areas with respect to domain closure, dialogue coherence and dialogue objective.

This section begins with a brief introduction into the history of artificial intelligence and conversational agents. After that, a selection of chatbots and information retrieval systems is presented. Finally, this section presents more relevant work from the tutoring domain and other areas.

#### 1.5.1 Artificial Intelligence

The topic of conversational agents is always closely related to artificial intelligence (AI). One of the most important milestones in the history of AI was the proposition of the Turing Test by the British mathematician Alan Turing in 1950 [Turing, 1950]. The intention of the Turing Test was to determine if a machine has thinking skills equivalent to humans. Turing suggested that if humans were unable to distinguish the computer's responses from that of a human, the computer could be said to be thinking and thus can be called intelligent.

The Turing Test had a strong influence on the definition of AI. Nowadays, literature distinguishes between strong and weak AI. A computer system, that is capable of full human

cognitive skills is called strong AI. Weak AI systems, in contrast, are limited to solving specific problems or reasoning tasks. However, the proposition of the Turing Test led to philosophical questions on strong AI, which are discussed until today. The most famous critic of the Turing Test was John Searle with his Chinese Room experiment [Searle, 1980] that should identify the Turing Test as an insufficient criterion for strong AI.

Regardless of the criticism on the Turing Test, Hugh Loebner established the first formal instantiation of a Turing Test in 1990, called the Loebner Prize <sup>3</sup>. The first computer system to pass the Turing Test is pledged \$100,000 and a gold medal and an annual contest awards the most human-like computer. However, in contrast to the large number of conversational agents participating in this competition, VPINO is not intended to pass the Turing-test. Needless to say, in a professional setting such as decision coaching, the client should always be aware that his partner is not a human.

### 1.5.2 Chatbots

While the existence of strong AI is still discussed, a large number of conversational agents have been developed. The first one was the famous chatbot ELIZA, presented by Joseph Weizenbaum in 1966 [Weizenbaum, 1966]. ELIZA is a simulation of a Rogerian psychotherapist, implemented using a rather simple pattern matching techniques. ELIZA's responses are generated by substitution of key words, which are extracted from the user's utterances, into predefined phrases. Apart from the implicit world knowledge hard-coded in ELIZA's transition rules, i.e. the pre-defined response phrases, ELIZA had no world-knowledge.

Although acting like a psychotherapist, ELIZA was not developed for that particular purpose. Moreover, it was rather intended as a technology demonstration than a therapist. Weizenbaum chose this scenario for rather simple, practical reasons: He argued that "(...) the psychiatric interview is one of the few examples of categorized dyadic natural language communication in which one of the participating pair is free to assume the pose of knowing almost nothing of the real world." [Weizenbaum, 1966]. Therefore, from a programming perspective, "(...) it eliminates the need of storing information about the real world." [Weizenbaum, 1966]. As a response to the user statement "I went for a long boat ride", ELIZA would respond with "Tell me about boats". Weizenbaum argued that one would not assume that the therapist knew nothing about boats, but that he had some purpose in directing the conversation in that direction. Surprisingly, ELIZA was taken seriously by a substantial number of users. Even after explaining the mechanism behind ELIZA, some of them were hard to convince of the fact that they have been talking to a machine. Nevertheless, with the role as a coach, VPINO also profits from this assumed credibility by the user.

Inspired by the "psychotherapist" ELIZA, Kenneth Colby introduced PARRY, a simulation of a paranoid patient during a session with his therapist [Colby, 1981]. Given the

---

<sup>3</sup><http://www.loebner.net/Prizef/loebner-prize.html>

transcripts of conversations, human psychiatrists were not able to distinguish between human patients and PARRY. In [Saygin et al., 2003] it is explained “Both ELIZA and PARRY use certain tricks to be able to successfully perform in conversations. ELIZA directs the conversation away from herself by asking questions. (...) In addition, PARRY has little Stories to tell and tends to insert these in the conversation”.

### 1.5.3 AIML and ALICE

Inspired by the success of ELIZA and similar chatbots, development resulted into the Artificial Intelligence Markup Language (AIML). AIML is a pattern matching markup language based on XML<sup>4</sup>. Like ELIZA, it makes use of a pattern matching approach and follows a simple “stimulus-response” principle. AIML basically consists of a large “knowledge collection”, a set of rules. These rules are represented as pattern-template pairs (rules):

- *Patterns* match the users’ text input with a subset of regular expression syntax. The pattern language consists of words, spaces and wildcard symbols.
- *Templates* define the AIML response in case of a pattern match. The simplest response for a matched pattern is plain text. Additionally, templates are able to set and read simple unary attributes and invoke other programs, e.g. call javascript. Templates can also recursively call the pattern matcher with a text defined inside the template.

Recursive substitution is probably the most important feature of AIML. It is used to simplify user language, normalize and correct spelling and grammar, detect keywords or synonyms, and transform user input into categories and concepts. Although very powerful, this feature of AIML is also considered its biggest weakness. The possibility writing recursively self-modifying scripts makes AIML systems hard to author and even harder to maintain. Tracing a user response or debugging is nearly impossible. Apart from the attributes, which are mainly used to store names or topics, AIML conversations are stateless. The attribute values cannot be accessed by the patterns. Also, AIML systems do not maintain and work with a conversation history.

A large number of the winners of the annual Loebner-contest in the last decade were based on AIML. The most famous chatbot implemented with AIML is A.L.I.C.E. by Richard S. Wallace [Wallace, 2009]. However, A.L.I.C.E. is a “classic” chatbot that aims to entertain users and pretend to be human. Its original version contains about 100,000 units of knowledge that were manually authored by a human. According to [Wallace, 2009], AIML implements a form of supervised learning where a person he calls “botmaster” plays a crucial role in training the bot.

---

<sup>4</sup>XML is a human readable markup language.



The open source chat interpreter ChatScript<sup>5</sup> was developed as an improvement of AIML. Instead of matching text only, it enables the definition of concepts, for example: `concept: ~fruit (apple banana raspberry mango)`. Thereby, it leads to a large reduction to the number of rules compared to AIML. The chatbot *Suzette*, which won the Loebner Prize in 2010, was written in ChatScript.

Other systems, like *Cleverbot*<sup>6</sup> or *MegaHAL*, learn their content from conversation with human users [Hutchens and Alder, 1998]. These unsupervised learning systems are crowd-sourcing the bot content to the users. Unsupervised learning has its own drawbacks, for example that the “(...) bot database becomes filled with nonsense, which then an editor must later delete. The trade-off between supervised and unsupervised methods might be summarized as ‘creative writing vs. deleting garbage’.” [Wallace, 2009, p. 206].

Although AIML based chatbots like A.L.I.C.E., but also user trained chatbots as MegaHAL, show a good response robustness for most of the user input on various topics, they were not intended for long, structured, coherent conversation on domain specific tasks, like decision support. Chatbots follow the “stimulus-response” principle at all times, and therefore lack intentionality. As the term “chatbot” suggests, the quality of a conversation with a chatbot can be characterized more as chat or smalltalk instead of holding a coherent conversation. Consequently, AIML is restricted to local dialogue coherence which makes insufficient for coaching conversation. The implementation of a coach requires to hold a coherent, goal oriented conversation over more than two consequent turns with the user.

Nevertheless, AIML is a widely spread amongst conversational agents used in practice. However, most practical application scenarios (e.g. information retrieval, see Sect. 1.5.7) do not require holding long structured conversation. Apart from entertaining chatbots, a lot of conversational agents have been implemented for practical application scenarios, many of them using well established AIML. [Shawar and Atwell, 2007] give an overview over a variety of chatbots for multiple purposes, with the focus on AIML based systems. However, most of the systems used in practice focus on providing domain specific information. For example, [Latorre-Navarro and Harris, 2014] present a rather simple AIML based assistant for university students that provides access to academic information based on a FAQ-like corpus.

### 1.5.4 Façade

Another notable conversational agent, but from a completely different domain, is the computer game Façade [Mateas and Stern, 2003]. Façade is an interactive drama that puts the player in a 3D environment together with two artificial characters, with whom the player interacts and talks (chats) in natural language. The story of the game develops by chatting with the artificial characters Will and Grace who strive about their marriage. The player

<sup>5</sup><http://sourceforge.net/projects/chatscript>

<sup>6</sup><http://www.cleverbot.com>

modifies the relationship to the characters, and their internal “emotional” states (e.g. increasing or decreasing the happiness value, by interacting, agreeing with, complimenting or insulting one of the artificial characters).

Façade is an interesting piece of work since its basic idea for their agent’s behaviour is similar to the one used with VPINO: Façade does not try to fully understand the input. Instead, the player’s text input is interpreted as discourse acts, a representation of the general meaning of the player’s action, similar to VPINO’s dialogue acts. The pattern matching language in Façade is slightly more complex than AIML. For example they extended the wildcard operator and allowed the OR and NOT operator. Although very powerful, Façade’s pattern matching language was criticised for being wordy and too hard to author, as also criticised on AIML.

However, a conversation with Façade lacks intentionality. A conversation in Façade, although structured, has no fixed goal or objective. However, conversation focuses on discourse, with potentially more than just two communication partners. Façade aims to create a dramatic, entertaining story with different possible endings.

### 1.5.5 Conversational Agents in Education

A notable example for natural language conversational agents from the educational domain is AutoTutor [Graesser et al., 1999, Graesser et al., 2004]. It is a text-based, mixed initiative system that simulates the discourse patterns and pedagogical strategies of a typical human tutor in an introductory computer literacy course. It is also equipped with a talking head embodiment, capable of performing gestures. Based on a set of topics from computer literacy domain, AutoTutor provides four different types of questions, also called topic formats: (1) Question - Answer, (2) Didactic information - Question - Answer, (3) Graphic display - Question - Answer, (4) Problem - Solution. Each topic format includes a main question that is presented to the learner.

To determine the type of the students reaction, responses are classified into speech acts. In contrast to VPINO, AutoTutor only distinguishes between five simple answer types, i.e. Assertion, W/H-question(Where, What, How etc), Yes/No-question, directive or short response to determine whether the system has to evaluate an assertion or to answer a question. In order to evaluate the quality of the students’ answers to reasoning questions, AutoTutor makes use of Latent Semantics Analysis [Landauer and Dumais, 1997], a technique for analysing relationships between semantic concepts in text documents. Based on the evaluation results, it also provides feedback on the students answers and possibly pumps for more information if needed.

### 1.5.6 Conversational Agents for Coaching

As already mentioned in the section on e-coaching (See Sect. 1.3.3), professional coaching has not been extensively used with conversational agents. In the healthcare domain, SimCoach is an animated virtual assistant with the focus on promoting access to specific health care information [Rizzo et al., 2011, Morbini et al., 2014]. SimCoach is an embodied, mixed initiative system. Unlike the name suggests, it does not actually coach the users, at least not in our sense of “coaching”. Moreover, SimCoach provides information and advice for veterans with the focus on health care problems (e.g. psychological health, traumatic brain injury, addiction, etc.). It therefore engages users in a conversation and encourages them to take the first step and seek help.

### 1.5.7 Information Retrieval / Question Answering Systems

Apart from the domain of chatbots, natural language interfaces are used in the domain of information retrieval (IR), personal assistance and question answering (QA). IR and QA systems aim to provide answers (either in natural language or as structured information) to user requests in natural language. IR systems make use of advanced natural language processing and machine learning techniques to analyse, extract and generate new semantic information based on given explicit knowledge base. This knowledge base is usually provided as a large set of unstructured information (i.e. a collection of texts). However, VPINO is not a QA/IR system. Therefore, we will not get into the detailed mechanisms of information retrieval systems.

With their state of the art information retrieval system *Watson* [Ferrucci et al., 2010], IBM introduced the term of *cognitive computing systems* as a way to “(...) learn and interact naturally with people to extend what either humans or machine could do on their own. (Cognitive computing systems) help human experts make better decisions by penetrating the complexity of Big Data.” [IBM, 2015] Although IBM’s definition of cognitive computing to “help humans make better decisions” reads similar to decision coaching with VPINO, there is one big difference: Cognitive computing, a data-centric approach, aims to support humans by extending their knowledge base for decision by processing “Big Data.” VPINO, on the other hand, supports rational decision-making with a client-centric approach. Although QAs use text (or speech) based natural language interfaces, they are not intended for holding a coherent conversation. An overview on QA/IR in restricted domains can be found in [Mollá and Vicedo, 2007]. A list of AIML based QA/IR systems are presented in [Shawar and Atwell, 2007].

## 1.6 Contributions and Findings

In this work, we explore the potential of information technology for simulated coaching using natural language. More specifically, we developed VPINO, a prototypical text based dialogue system in German language. We implemented a rational decision coach and a training transfer coach for communication skills. Our artificial coach intends to support clients with reflection on their goal accomplishment process. We apply the technique of *Socratic questioning*. Instead of suggesting or instructing the client, we ask targeted questions. These coaching questions, as well as all other possible responses by the system, are precisely preformulated. By keeping the leading role and guiding the client, the system is able to keep track of the conversation context to a reasonable extent. Thereby, our dialogue system does not need to fully understand the semantic content of the client's responses in detail. Instead, the system classifies the client's reactions as *dialogue acts*. With a model for common dialogue act sequences, VPINO keeps track of the state and direction of the conversation.

Of course, the goal of the work presented in this thesis is not to replace real human coaches in general; just like Joseph Weizenbaum was sure that his famous chatbot ELIZA is not a way to replace psychotherapists. Our vision is a low threshold computer-based transfer coaching for situations where a personal human coach is simply not affordable or available. This thesis is a first step in that direction.

We conducted a series of studies to examine whether humans would accept such a system and consider it a useful tool. We evaluated our tool in a proof of concept and applied it to the concrete coaching scenarios *training transfer coaching* and *rational decision coaching*.

### 1.6.1 Proof of Concept

In order to find out whether our approach is applicable on professional coaching in general, we conducted a pilot study and tested a prototype as a proof of concept. We implemented a general Socratic coach on 10 students (See Chapt. 3). Subject of the conversations were career related problems and decisions. The results of this pre-study convinced us that our approach is promising for more specific professional tasks such as rational decision coaching. The study also revealed the necessity for more human-like conversations and grounding techniques (See 2.4). In addition, these first results suggested that the clients personality and attitude have an impact on the quality and efficiency of our tool.

### 1.6.2 Scenario I: Training Transfer

We used our dialogue system as a training transfer coach for communication skills. We conducted a user study where the clients received an online communication training on

weekend followed by a week of daily coaching sessions with our dialogue system. The clients daily work and progress on their communication skills were the topic of the conversations. Our computer based coaching was tested against a conventional transfer method with respect to effectiveness and influence of the users personality and attitude. As a result of the study, we found that by using our tool clients could successfully improve their communication skills. In accordance with our first qualitative impressions from the pre-study, we found that the personality trait *openness* of the client has an impact on the outcome of the transfer coaching: More open users that received a daily coaching intervention by our system particularly improved their communication skills. Parts of the results of this user study have been published in [Mäurer et al., 2013].

### 1.6.3 Scenario II: Rational Decision-Making

The second application scenario for our dialogue system is *rational decision coaching*. We also implemented a version of VPINO specifically tailored to support humans making decisions on a distinctly rational basis. The method for rational decision-making is an adoption of Benjamin Franklin's famous *Pros and Cons*. Franklin's approach was extended with more modern concepts such as goal setting. Our dialogue system employs a client-centric coaching approach where the client is seen as the real domain expert. Therefore, the system does not make a decision for the client, nor does it suggest a particular option. Instead, the dialogue system guides and leads the user through the chat conversation while hiding the details of the underlying methods on decision-making.

#### User Study on Decision-Making

We conducted a user study to measure the clients goal attainment when using our coach, and factors that moderate success of the conversation. The participants were free to choose a rational decision problem of their choice and speak to the coach about that problem. Overall, the decision coach was evaluated quite positively. We were able to help a large number of participants, either with clarification of their situation or guidance on making a decision. The results suggest that the clients personality, motivation, cooperation and their usual approach to decisions affect success of the conversation. Participants with a rather hypervigilant approach on making decisions profit in particular from holding a conversation with our dialogue system. However, a successful conversation requires a reasonable level of motivation and cooperation from the user. Clients who do not take the system seriously enough or try to challenge it will not enjoy target-aimed dialogues. Parts of the work on rational decision coaching with a conversational agent, including the results of this user study is published in [Mäurer and Weihe, 2015].

### **Follow-up User Study on Decision-Making**

We further improved our decision-making dialogue plan based on the insights from the first study. Previous results suggested that showing off intelligent, human like behaviour, motivate participants for more cooperative behaviour. Therefore, we improved several components related to intelligent behaviour, e.g. object reference resolution for particular options or arguments and the system's capability on pairwise option comparison. We conducted a follow up study with a larger, more representative group of participants. In general, the results of the first study could be confirmed. VPINO was able to reduce the level of emotional stress for more hypervigilant decision makers. Qualitative analysis of the transcripts revealed that the best human like conversations were produced by older participants, whereas most of the younger participants tend to give short answers with a few words only. Whereas VPINO was effective for more rational problems, it seemed to perform not perfectly for emotional problems.

## CHAPTER 2

# Background and Concepts Behind VPINO

---

This chapter intends to give the reader detailed insight of the concepts and mechanisms behind the VPINO dialogue system. Well established technology for chatbots, such as AIML, does not fit the requirements for structured, coherent coaching conversation. Conclusively, we have developed a conversational agent specifically for computer based personal coaching conversation, based on the principles presented in the previous chapter.

**This chapter is structured as follows:** Sect. 2.1 provides an overview of the basic mechanisms behind VPINO. The theory of “Language as action”, i.e. speech act theory or dialogue act theory, will be presented in Sect. 2.2, followed by a description of the dialogue act classification component in Sect. 2.3. In Sect. 2.5, we introduce sub-dialogues and VPINO’s response planning mechanism. The chapter is closed with a detailed description of VPINO’s turn-taking and communication concept in Sect. 2.6.

## 2.1 Big Picture of Vpino's Background

The response mechanism of VPINO differs largely from the common pattern matching approach. The main differences to AIML-like pattern matching are a the turn-taking mechanism, a structured knowledge memory, and the use of local sub-dialogues for response planning, based on an advanced pattern matching language for dialogue act classification. For a human-like user experience while talking to VPINO, which is necessary for a constructive relationship with the client, VPINO follows well studied concepts of communication theory, such as the establishment of a common ground and human turn-taking behaviour.

Turn taking, i.e. choosing the proper timing to contribute an utterance (a turn), is essential in human conversation [Sacks et al., 1974]. VPINO implements an asynchronous, mixed initiative turn-taking protocol. In contrast to the simple “stimulus-response” mechanism, users are free to contribute their turn at any time of the conversation with VPINO. For more human-like behaviour, VPINO also simulates the physical act of “typing a message” by delaying the contribution of its turn instead of simply responding to user immediately.

In order to hold a structured, goal-directed conversation, VPINO keeps an internal *structured knowledge memory (SKM)*. The SKM is a flexible, nested hash data-structure to store all data/information collected throughout the conversation. The data stored in the SKM can be accessed from any other component in the system, except the dialogue act classification rules, which do not need access to the internal memory.

Natural language understanding of VPINO is based on the *dialogue act* of a user utterance. The dialogue act of an utterance is the communicative function performed by the speaker [Austin, 1962, Searle, 1969]. Instead of directly mapping the user input to a response, VPINO determines the dialogue act the user utterance performs. The system’s actual response depends on that dialogue act, the current context of the conversation, and the data stored in the SKM.

Conversation with VPINO is organized as a collection of sub-dialogues. A *sub-dialogue* is a model for all potential sequences of turns in an enclosed local context. The context of a particular sub-dialogue is defined by a corresponding coaching question. Each sub-dialogue targets a specific piece of information that is extracted and stored in the SKM. A sub-dialogue manages conversation in three phases:

1. In an short *grounding* phase, prior to asking the particular question, VPINO prepares the client for the topic of the upcoming sub-dialogue. Grounding theory describes the presupposition and accumulation of a *common ground*, i.e. an initial set of mutual knowledge and beliefs by the conversation participants that is added up throughout the conversation [Stalnaker, 1978].
2. The following *dialogue act sequence phase*, which handles the actual conversation, is organized as a finite-state-machine like model. The model represents all expected sequences of dialogue acts to answer the corresponding coaching question. De-



pending on the state of the sub-dialogue, determined by the sequence of dialogue acts by the user and VPINO so far, the system selects a response. Response texts are pre-authored for each particular sub-dialogue. However, responses can use place-holders in order to refer to pieces of information stored in the SKM.

3. In the data evaluation phase, all user utterances in the current sub-dialogue are parsed for relevant information, as targeted by the coaching question. The extracted data is evaluated, and the SKM is finally updated with the new piece of information.

After finishing a sub-dialogue, the following sub-dialogue is selected based on the state of the SKM and the *dialogue plan*. The dialogue plan is a manually authored broad plan for the whole coaching conversation. It includes manually authored precedence constraints for particular sub-dialogues, and provides control structures for loops and if-conditions.

## 2.2 Dialogue Act Theory

VPINO's natural language understanding capability is based on dialogue act sequences and dialogue act classification. This section provides background information on dialogue act theory in general and the dialogue act classification component implemented for VPINO in particular.

### 2.2.1 Background on Dialogue Acts

Communication theory describes conversation from a perspective of "dialogue as a sequence of joint actions" [Schegloff, 1990]. The basic idea for this perspective originates from Ludwig Wittgenstein's *ordinary language philosophy*. Wittgenstein claims that the meaning or function of language depends on its actual use in a context, rather than having an inherent meaning. By speaking out, the speaker is performing a context specific action. The action performed by an utterance, or the function of an utterance, is the subject of *speech act theory* and *dialogue acts theory*.

**Speech Act Theory** Speech act theory was first introduced by John L. Austin [Austin, 1962]. The basic idea of speech act theory is to describe the consequences and effects of an utterance. The consequences of an utterance are not limited to conversation in which it is uttered. In fact, these consequences have real effect on the world. For example, by speaking out the phrase "*I now pronounce you husband and wife*", the speaker is performing the act of marrying a couple.

In his famous work "How to do things with Words" [Austin, 1962], Austin claims that each utterance by a speaker is constituted by its *locutionary*, *illocutionary*, and *perlocutionary*

act. The locutionary act is the physical act of uttering a sentence. The illocutionary act is the action performed by an utterance (i.e. asking, bidding, agreeing, etc.). The illocutionary act is also referred to as the *speech act* of that utterance. The perlocutionary act is the intended effect of an utterance on the world (or at least the receiving addressee). For the previous example “*I now pronounce you husband and wife*”, the locutionary act is the utterance of speaking out a declaration, the illocutionary act is the act of declaring, and the perlocutionary act is changing the legal status of two persons in the real world.

John R. Searle, one of Austin’s followers, refined Austin’s speech act theory with a taxonomy of illocutionary acts [Searle, 1969]. Searle introduced the five major classes as a first speech acts taxonomy: Assertives aim to describe a state of affairs in the world and thereby commit a speaker to the truth of the expressed proposition. Directives intend the receiver to take action, e.g. requests, commands and advice. Commissives commit a speaker to future action, e.g. promises and oaths. Expressives express emotions towards the proposition, e.g. congratulations or thanks. Declarations change the reality/state of the world in accord with the proposition of the declaration, e.g. naming someone, pronouncing someone guilty or pronouncing someone husband and wife.

**Dialog Act Theory** The term “dialogue act” can be described as a generic term for specialized speech act. Unlike for speech act theory, there is no common definition for dialogue acts in general, or a particular dialogue act taxonomy. Based on Austin’s and Searle’s speech act theory, several theories and taxonomies like *communicative acts* [Sadek, 1991, Allwood, 1976], *conversation acts* [Traum and Hinkelman, 1992], *conversational moves* [Carletta et al., 1997], and *dialogue moves* [Traum et al., 1999] have been presented. However, all of them describe similar concepts and will be therefore referred to as *dialogue acts*.

A popular taxonomy for dialogue acts is the *Dialogue Act Markup in Several Layers* (DAMSL) scheme by [Core and Allen, 1997, Allen and Core, 1997]. The DAMSL taxonomy provides a high-level description of communicative actions with multiple layers. In addition to the basic definition of speech acts, the taxonomy also includes the notion of dialogue structure.

The Switchboard corpus a collection of human-to-human telephone conversations [Godfrey et al., 1992], defines 42 types of dialogue acts (SWBD-DAMSL). Table 2.1 provides an overview on the dialogue act tags and their frequencies in the Switchboard corpus.

Although spoken conversation on general topics in the SBWD-DAMSL differs from the scenario of text-based chat coaching in many ways, frequencies reveal the most common dialogue acts in conversation: Apart from *Statements*, which are the most common dialogue acts, other frequent dialogue acts are related to a communicative function instead of semantic information: For example, 19% of the dialogue acts are tagged as Acknowledge (Backchannel) acts, which have the purpose of grounding (See Sect. 2.4). However, the frequencies of dialogue act occurrences largely depend on the type of conversation in the

Dialogue Act	Example	Freq
Statement-non-opinion	Me, I'm in the legal department.	36%
Acknowledge (Backchannel)	Uh-huh.	19%
Statement-opinion	I think it's great	13%
Agree/Accept	That's exactly it.	5%
Abandoned or Turn-Exit	So, -	5%
Appreciation	I can imagine.	2%
Yes-No-Question	Do you have to have any special training?	2%
Non-verbal	[Laughter], [Throat_clearing]	2%
Yes answers	Yes.	1%
Conventional-closing	Well, it's been nice talking to you.	1%
Uninterpretable	But, uh, yeah	1%
Wh-Question	Well, how old are you?	1%
No answers	No.	1%
Response Acknowledgement	Oh, okay.	1%
Hedge	I don't know if I'm making any sense or not.	1%
Declarative Yes-No-Question	So you can afford to get a house?	1%
Other	Well give me a break, you know.	1%
Backchannel in question form	Is that right?	1%

**Table 2.1:** Frequencies of dialogue acts in SWBD-DAMSL corpus. The table displays only dialogue acts with frequencies  $\geq 1\%$ . Sum of remaining dialogue acts make up  $\approx 3\%$ , e.g. *Quotation*, *Rhetorical-Questions*, *3rd-party-talk*, *Apology* and more.

corpus. For example, conversations in the tutoring domain include much different types of questions.

### 2.2.2 Taxonomies for Chat Conversation

The DAMSL taxonomy is a complex annotation scheme for theoretical conversation analysis and is not intended as a tag set for the development of a conversational agent. For practical use in a dialogue system, a large number of tags are not relevant. The majority of conversational agents either does not determine the dialogue act of user utterances at all, or is using only a very small set of dialogue acts. That is due to the simple fact that most of these systems are based on AIML or other pattern matching approaches, where a match is directly mapped to a response. Thereby, the perspective of conversation as a sequence of dialogue acts is irrelevant.

However, there are few conversational agents that make use of dialogue acts for response planning: An exception makes the computer game *Faade* [Mateas and Stern, 2003] by mapping user input to *discourse acts* ( $\approx$  dialogue acts). Conversation in *Faade* focuses on social interaction in a dramatic, tense, and emotional discourse. Consequently, their set of discourse acts mirrors this scenario: The majority of their discourse acts describes social interaction, e.g. *Apologize*, *Praise*, *Criticize*, *Flirt*, *Pacify*, *Ally*, *Oppose* and alike. AutoTutor [Graesser et al., 2005] also classifies dialogue acts, but only distinguishes between five types: *Assertion*, *W/H-question*, *Yes/No-question*, *Directive* or *Short response*. [Schulman et al., 2011] use a domain dialogue acts for healthcare related motivational interviewing. However, they do not describe their taxonomy in detail.

According to [Traum, 2000], specific deficits in any given theory often lead researchers to continue to develop new taxonomies to suit their particular purposes. Thus, dialogue act taxonomies can be seen as a kind of language for describing communicative events for a particular purpose. According to [Popescu-Belis, 2003, p. 3], the definition of a dialogue acts set should be “(...) motivated by looking at the functions of actual utterances in a given domain” and “(...) should mark the functions that are important instead of trying to mark all functions”. Besides, they suggest the tag set not be too remote from the present capabilities of NLP. In order to build a model for a structured coaching conversation, we have developed a set of dialogue act specifically tailor for our purposes, instead of reusing an improper one developed for a different domain.

### 2.2.3 Dialogue Acts Taxonomy for Structured Coaching

In order to determine a suitable set of dialogue acts for structured Socratic coaching, we conducted a *Wizard-of-Oz* (WOZ) experiment. WOZ is a research experiment technique in which human subjects interact with a computer system, that they believe to be autonomous, but is actually being operated or partially operated by an unseen human op-

erator [Hanington and Martin, 2012]. WOZ experiments are carried out to get a first idea of potential user reactions in the development of dialogue systems.

With an initial set of essential coaching questions, taken from literature on different coaching approaches, the simulated coach responded to the users answers without understanding semantic details of the users response. Instead, the system followed the *cooperative principle* [Grice, 1970] and the concept of *adjacency pairs* [Schegloff, 1968]. The *cooperative principle* is a description of how humans normally behave in conversation and consists of four maxims, also known as *Grice's Maxims*:

1. Maxim of quantity: Make your contribution as informative as required. Don't make your contribution more informative than is required.
2. Maxim of quality: Don't say what you believe to be false.
3. Maxim of relation: Be relevant.
4. Maxim of manner: Be perspicuous. Avoid obscurity of expression and ambiguity. Be brief and orderly.

An *adjacency pair* is a pair of turns by two speakers, where the act performed by a speaker in the first turn is provoking a valid response in return. A response is valid, if it does not violate the maxims of conversation, e.g. it is not relevant. Contribution an irrelevant response, i.e. a fail to complete the pair, is a breach of conversational maxim. Examples for adjacency pairs are *greeting* → *greeting*, *offer* → *acceptance/rejection*, *compliment* → *acceptance/refusal*, or *question* → *answer*.

Participants were instructed to respond naturally, just as they would when talking to a human. Thereby, we collected a set of conversation snippets for each particular coaching-question. For these conversation snippets, we identified the most common adjacency pairs of dialogue acts and short dialogue act sequences. With the experiment, we were also able to identify more and less suited questions for computer based coaching with VPINO, for example misleading or too open questions. Based on these simulated coaching dialogues, we implemented a simple baseline classifier for the most common dialogue acts and conducted a pre-study, which will be presented in Ch. 3. With the resulting conversation transcripts from the pre-study and the dialogue act set collected in the WOZ experiment, we determined a set of dialogue acts that are suitable for structured coaching conversation with VPINO. The list of dialogue acts is a subset of the common DASML tags, and is presented in Tab. 2.2. Note that we decided not to split up different types of statements, since they were often semantically related to the preceding coaching question and will therefore be evaluated separately. The dialogue act sequences and adjacency pairs collected in the study were also used for the construction of sub-dialogue models, which will be described in Sect. 2.5.

Dialogue Act	Example Text
WH-Question	What do you mean with ...
YN-Question	Did you work on your goals today?
Other-Question	Mhm?
Statement	I have been working all day.
Acknowledgement/Backchannel	mhm okay. . .
Agreement/Positive answer	Yep, sure it is.
Disagreement/Negative answer	I don't think that's right.
Pass	not really sure about this...
Clarify	I mean the exact amount of money to buy the BMW.
Insist	I need an answer, otherwise we will not continue
Pump	Tell me more about this.
Opening	Hi there!
Closing	See you!

**Table 2.2:** The set of dialogue acts used with VPINO

### 2.3 Dialogue Act Classification with Vpino

Dialogue act classification is a task of information extraction. Approaches to information extraction from text can be roughly divided into *rule-based* and *machine learning* approaches. Whereas rule-based approaches are authored by domain experts, machine learning approaches use probabilistic methods to train a model from labeled data. However, there are advantages and disadvantages for both. Rule-based approaches are declarative and readable, and therefore easier to comprehend and maintain. Nevertheless, a large amount of manual effort is required for authoring a good set of rules. Machine-learning approaches, on the contrary, reduce manual effort with adaptive, trainable models. On the other hand, machine learning requires a set of labeled, domain specific training data. For the domain of coaching or counselling, there are neither plain transcriptions, nor tagged data for chat coaching in German language due to privacy restrictions in coaching conversation. Whereas probabilistic machine learning approaches dominate in research and academia, commercial systems are still dominated by rule-based approaches [Chiticariu et al., 2013]. The requirements of real-world use cases often lead to the use of rule-based approaches, mostly due to availability of domain specific tagged corpora, and the traceability of classification results. In the particular domain of chatbots and dialogue systems, pattern matching approaches based on regular expressions are still very popular (See Sect. 1.5).

Nevertheless, a number of academic research on machine-learning approaches to dialogue act classification has been presented [Stolcke et al., 2000, Marineau et al., 2000, Reithinger and Klesen, 1997, Kim et al., 2010, O'Shea et al., 2012]. However, there are of course no clear boundaries between rule based and machine learning approaches. Hybrid information extraction systems can apply components of both approaches are a good choice for real-world use cases.

### 2.3.1 VPINO Classification Pipeline

VPINO uses a rule-based dialogue act classifier. The set of classification rules is applied to pre-processed text. More specifically, input text is annotated with syntactic information from probabilistic models trained on large corpora. The rule language is provided by UIMA Ruta<sup>1</sup> [Kluegl et al., 2015]. Ruta is a component for the Apache UIMA framework [Ferrucci and Lally, 2004]. UIMA stands for Unstructured Information Management Architecture. It is a software framework implementation for the analysis of unstructured data, like text. The UIMA framework specifies component interfaces in an analytics *pipeline* for natural language processing (NLP). The pipeline inflicts specific analytic components to annotate an input text. An *annotation* marks a span in the input text with a type and optional meta-information. The annotated text object is passed to the next component in the pipeline. Output of an analytics pipeline is the input text document with annotated meta information, e.g. the dialogue act performed by uttering the input text. VPINO's text classification pipeline is built with the DKPro toolkit [de Castilho and Gurevych, 2009], a collection of natural language processing components for the UIMA Framework.

VPINO implements a classification pipeline as displayed in Fig. 2.1. It consists of a tokenizer, a POS-tagger, a lemmatizer and a set of dialogue act classification rules. Input of the pipeline is the plain text of a user utterance, the output is an annotated object including the dialogue act of the text. Although UIMA is originally intended for batch processing large collections of input documents, we have implemented a wrapping component to provide dialogue act classification as a RESTful web-service.

### 2.3.2 Preprocessing Components

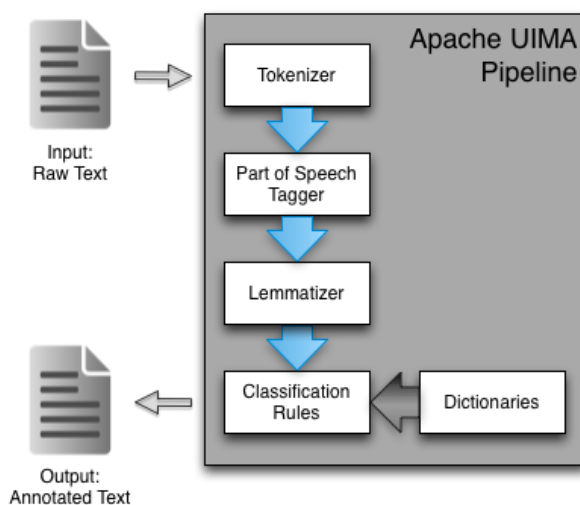
The following components in the classification are used for preprocessing the input text before application of the classification rules:

**Tokenizer** Tokenization is the process of splitting a text into a set of tokens. Tokens are smaller units of text, for example sentences, words, phrases or symbols. Tokenization is also referred to as word segmentation. Tokenized/segmented text is a required by further components in the pipeline.

**Part-of-Speech (POS) Tagger** *Part of speech*, also known as *word class*, *lexical class*, or *lexical category*, describes groups of words with similar roles in grammatical structure of a sentence. Examples for part of speech are noun, verb, adjective, adverb, preposition or pronoun. For each word of a given input text, the POS-tagger annotates the part of speech that word.

---

<sup>1</sup><http://uima.apache.org/ruta>



**Figure 2.1:** Schematic diagram for the VPINO dialogue act classification pipeline. Input text is split into words and sentences, annotated with lemmas and pos-tags before a set of classification rules determine the dialogue act of the input text. Annotated text is returned

**Lemmatizer** In written language in general, words mostly appear in several inflected forms. A lemma is the canonical form of a group of words. A *lemmatizer* is a component that determines the base form of a word in any inflected form. For example, the conjugations of the words *be*, *am*, *was*, *been* share the lemma “*be*”. Lemmatization is often used for normalising irregular conjugations. Normalized text makes the task of authoring a set of rules a lot less extensive.

The classification pipeline implemented for VPINO uses TreeTagger, a probabilistic approach for POS-tagging and Lemmatization developed by Helmut Schmid [Schmid, 1994]. TreeTagger was chosen since it is considered to work particularly well for German language.

### 2.3.3 Rule-Based Text Annotation with Ruta

Finally, the preprocessed text is annotated by a set of classification rules using UIMA Ruta [Kluegl et al., 2015]. Ruta stands for Rule-based Text Annotation and is a component for the UIMA framework. Ruta provides an imperative rule language for matching and annotating text. As a part of the UIMA framework, Ruta was developed for batch processing large collections of documents and complex NLP tasks. Whereas AIML was developed for the purpose of matching short text in the particular context of chat conversation, Ruta allows more complex rules. Nevertheless, it can be used for very short text as well.

The Ruta language has a number of advantages over reg-exp pattern matching: It allows the definition of expressions and variables, score-based filtering and flexible control



structures. Also, it supports modularization in different files or blocks and the invocation of dictionaries, which is a huge organisational advantage for authoring and maintaining rules.

Ruta matches on abstract concepts, i.e. types of annotations. New types of annotations can be defined and reused in other rules. In contrast, regular expressions (or AIML) match on text patterns only. Also, Ruta's language definition allows to import annotations from external components as well. Thereby, rules can access the annotations from the pre-processing components, such as the segmented words, sentences, POS-tags, and lemmas.

### 2.3.4 The Ruta Rule Language

In the following, we give a brief overview on the Ruta rule language. However, this thesis will not provide a detailed description of the Ruta language. A comprehensive documentation can be found on the official UIMA Ruta website<sup>2</sup>.

A program with Ruta basically consists of a script, which contains a collection of rules. Rules define a matching condition and modify a matched text by adding/editing meta-information, i.e. annotating particular parts of the matched text. More precisely, the rules are used to create new annotations corresponding to the dialogue act of a text. A Ruta script is defined as follows: A script is a collection of rules. A rule is composed of a sequence of rule elements. A rule element essentially consists of four parts: A *matching condition*, an optional *quantifier*, a list of *additional conditions* and a *list of actions*.

- The **matching condition** is a rule element that matches on the text covered by the annotation type defined in the matching condition. Besides matching on a type of annotation (e.g. a sub-string annotated with the type "Noun"), a matching condition can also match strings (e.g. the plain word "Noun").
- The **quantifier** specifies whether it is necessary that the rule element successfully matches and how often the rule element may match (e.g. zero or more, one or more, zero or one time, exactly x times, ...). The definition of quantifiers is similar to those used in regular expressions (\*, +, ?).
- The **list of additional conditions** specifies additional constraints that the matched text or annotations need to fulfil to perform the actions defined in the list of actions. Besides basic logical functions, theses additional conditions can evaluate a number of predefined functions (e.g. STARTSWITH, COUNT, NEAR, ...).
- The **list of actions** defines the consequences of the rule in case of a match. This consequence is usually creating a new annotations or modifying/removing existing annotations.

---

<sup>2</sup><http://uima.apache.org/ruta>

### 2.3.5 Example Dialogue Act Classification Rules

The following rules are examples for dialogue act classification rules implemented for VPINO. The examples are only a very small collection taken from the set of rules used with VPINO. However, they do not represent all rules for particular dialogue acts. The set of examples was chosen to explain the general dialogue classification mechanism with Ruta.

Note that functions of the Ruta language are written in CAPITAL letters. ‘‘Quoted text’’ denotes a plain string. Non quoted Terms denote any kind of annotations.

#### Example Wh-Questions

The following rule intends to detect and annotate ‘‘Wh-questions’’, which is a rather simple task and therefore making up a good first example:

```
1 Sentence{AND(STARTSWITH(INLIST(wq)),LAST(Question))->MARK(WQuestion)};
```

The *matching condition* of the rule matches on every Sentence in a given input text. More precisely, the rule element matches a range of characters, annotated as a Sentence by the tokenizer. In this example, the rule element does not specify a quantifier. Therefore, the condition will match every Sentence in the input text. A matched Sentence is annotated as a WQuestion, if the following additional conditions apply: The Sentence starts with a word listed in a dictionary of w-question words (i.e. a list of words including ‘‘where, what, who,...’’), and the Sentence closes with a question mark (i.e. the LAST element of the Sentence is annotated as Question).

#### Example Disagreement

The following example intends to annotate short sentences with a leading ‘‘no’’ as a Disagreement.

```
1 ‘‘no’’ W[1,4] Sentenceend{->MARK(Disagreement,1,2,3)}
```

Any span of text starting with a ‘‘no’’, followed by one up to four words W[1,4] and a Sentenceend annotation (i.e. period, question or exclamation mark) is marked as disagreement. The numerals in the MARK action indicate, that all of the three rule elements should be annotated as a Disagreement, and not only the Sentenceend. However, although the example rule might work for some Disagreements, it is poorly designed.

One advantage of Ruta over simple pattern matching is the capability of defining concepts and include dictionaries. The following set of rules makes use of that advantages by defining and reusing the concept of a Shortsentence and defining the concept of No-words based on a dictionary:

## 2.3. Dialogue Act Classification with Vpino

```
1 Sentence{COUNT(W,1,5)->MARK(Shortsentence)};;  
2 W{INLIST(no-words)->MARK(No)};;  
3 Shortsentence{STARTSWITH(No) -> MARK(Disagreement)};;
```

The first rule will annotate each *Sentence* as a *Shortsentence*, if the additional constraints *COUNT(W,1,5)* evaluates true, i.e. the number of word annotations *w* within the range of the *Sentence* is between 1 and 5. If so, the *MARK* function annotates the whole matching element (i.e. the *Sentence*) as a *Shortsentence*. The second rule annotates each word listed in the "no-words" dictionary *No* tag. In the third rule, each short sentence that starts with a part that is annotated as a *No* tag, is tagged a *Disagreement*.

### Example Passing

The rules in the following example are written to determine *Passing* acts by a user, i.e. answers like "I don't know". However, not every utterance where a user does not know something is a *Passing*. Therefore, the rule below marks a sequence of annotation types as a hint for the act of "Passing" if the following conditions are true: The first element is any representation of "I", followed by any form of verb matched as a synonym for "to know", and a negation.

```
1 I Negation Know {->MARK(PassHint,1,2,3)};;
```

A lot of work has been done in the pre-processing pipeline: The POS-tagger has identified a token as some form of the personal pronoun "I" and a *Negation*. The concept of the *Know* tag has been established by an additional rule:

```
1 V{INLIST(knowlist, Lemma.value)}{-> MARK(Know)};
```

The rule annotates a verb as *Know*, if the lemma (also annotated by the POS-Tagger) is listed in a dictionary of synonyms for "know". The following rules further extend the previous example:

```
1 I W[0,4]?{-CONTAINS(Negation)} Negation Know{->MARK(PassHint,1,2,3,4)};;  
2 PassHint{->UNMARK(PassHint)} COMMA? If;
```

The first rule allows up to four optional non negation words between the "I" and "not knowing" to match more complex utterances such as the German: "Ich weiß jetzt leider noch nicht genau wozu ich tendiere..." The second rule applies to text such as: "I am not sure whether i should go out or stay at home", which is clearly rather a *Statement* than a *Passing*. A *PassHint*, followed by an optional *COMMA* (obligatory for if in German grammar, but often skipped in text-chat) and some kind of "if" or "whether", is unmarked as a hint for a *Passing*.

## 2.4 Grounding Background

For a computer based chat system, the establishment of *common ground* with the user is a key success factor. Not only to make the conversation feel more human-like, but as an actual necessity for successful conversation with a human. This section will provide background on grounding theory and discuss methods for grounding with VPINO.

### 2.4.1 Grounding Theory

The accumulation of mutual beliefs or common ground by the speakers holding a conversation is a central part in conversation analysis [Cahn and Brennan, 1999, Paek and Horvitz, 1999, Traum, 1999, Brennan, 1998]. The term *common ground* was first described by Robert Stalnaker as “the common set of mutual knowledge that is taken for granted as the background of the conversation” [Stalnaker, 1978]. The accumulation of this knowledge is called *grounding*. The concept of *grounding*, the process of establishing mutual knowledge, beliefs and assumptions, was proposed by Clark and Brennan [Clark and Brennan, 1991]. However, common ground cannot be reached by one speaker alone. Both communication participants are needed to collaboratively establish this common ground.

The lack of a common ground in conversation has a huge impact: “When communicating partners fail to highlight the important points of their message to their partner or know the important points of the partner’s message, then both parties can never satisfy the their partner’s expectations (...) This lack of common ground damages interpersonal trust, especially when the communication partners do not share the contextual information of why the other party behaves the way they did.” [Cramton, 2001].

Reaching a common ground with a human user is a challenge for a conversational agent. Clark and Brennan identify eight characteristic constraints for communicating parties, namely *copresence (collocation)*, *visibility*, *audibility*, *contemporaneity*, *simultaneity*, *sequentiality*, *reviewability* and *revisibility*. Text based chat communication largely differs from spoken conversation in human face-to-face coaching, e.g. by a lack of eye contact, mimics, gestures, and prosody. Also, conversation analysis presupposes a common ground between both communication partners prior to starting the conversation. For communication between a computer system and a human user, such common knowledge or beliefs do not exist. User expectation is subjective and are mainly moderated by personal experience, either with human coaching or talking to computers in general. A lack of common ground leads to mutual misunderstanding, with the consequence of disappointment by both communication partners. The establishment and maintenance of common ground is essential for successful communication between VPINO and the client.

Clark and Brennan present a model that allows track the augmentation of common ground throughout the conversation [Clark and Brennan, 1991]. Their model distinguishes between the *presentation phase* and the *acceptance phase*. In the *presentation phase*, the speaker

presents an utterance. In the *acceptance phase*, the addressee shows acceptance of the presented utterance by providing evidence of understanding. However, the display of acceptance is not necessarily restricted to the use of language. For example, continued attention or nodding are also ways to show an utterance was received and understood.

The concept of presentation and acceptance are closely related to the perspective of “language as a sequence of action”: Presenting informations, showing understanding, agreeing or disagreeing on something can also be described as specific dialogue acts performed by a speaker and the receiver.

### 2.4.2 Methods for Grounding

This section presents a set of methods to reach common ground and evaluates their applicability for a chat based conversational agent. As mentioned before, VPINO is a mixed initiative system. Thus, in contrast to “stimulus-response” systems, VPINO should be able to contribute to common ground in both *presentation*, and *acceptance* phase.

Contribution in the acceptance phase is a more complex task, especially for an artificial speaker such as VPINO. In the acceptance phase, the task is to display that a message has been received and understood, which is crucial to general communication. [Clark and Schaefer, 1989] describe five main methods to display understanding in the acceptance phase:

1. Continued Attention: The addressee is satisfied with the speaker’s presentation. Instead of interrupting the speaker, the addressee shows acceptance by continuing to attend.
2. Next Relevant Contribution: The addressee confirms grounding by starting in on a next relevant contribution.
3. Acknowledgement: The addressee utters an acknowledgement or a continuer such as ‘yeah’, ‘uh-huh’ or the like to express reception and understanding.
4. Display understanding: The addressee expresses understanding by displaying verbatim of the speakers utterance, i.e. fully or partially repeating what the speaker just said.
5. Demonstrate understanding: The addressee demonstrates understanding by reformulating, paraphrasing, collaborative completion of the speakers turn.

Continued attention is considered a weak method for grounding, Display and Demonstration are considered the strongest methods to show understanding in conversation.

In communication settings without visual contact, such as online chat, continued attention is likely to provoke misunderstandings. When working with computer systems in

general, humans tend to interpret a lack of feedback as erroneous behaviour. As a result, continued attention is not a suitable method for grounding with a conversational agent. Keeping control of the conversation by *next relevant contribution* is a better way to establish common ground. However, timing the next contribution requires a suitable turn-taking mechanism (See Sect. 2.6).

Although the use of *acknowledgement*, *continuer* or *back-channel* as a grounding method is more frequent in spoken than in chat conversation, it is also crucial for a computer based system. The frequent utterance of acknowledgements is a simple, but effective way to cope with the problems of missing non-verbal feedback. A lack of acknowledgement is known to “(...) cause errors and confuse humans when talking to virtual agents” [Yankelovich et al., 1995]. The use of acknowledgement or continuer might seem an unimportant detail. Nevertheless, they have a great impact on successful human-like conversation.

Repetition of an utterance is considered a strong method of grounding in human conversation. For a computer coach, repetition does not automatically prove that a message was also “understood”. Apart from rather “simple” acknowledgements as a display of understanding, humans often demand a proof of understanding. Demonstrating understanding by paraphrasing or reformulating a speaker’s utterance requires detailed semantic knowledge about the utterance. Therefore, sub-dialogues include particular grounding phases for both presentation, and and acceptance phase, as will be presented in the following section.

## 2.5 Sub-Dialogues

In order to keep control of the conversation and contribute to common ground, VPINO is organised as a collection of sub-dialogues. Sub-dialogues are the building blocks of the VPINO dialogue system. As mentioned before, each sub-dialogue is devoted to answering a specific coaching question. The goal of a sub-dialogue is to retrieve a specific piece of information targeted by that particular coaching question. The targeted information is extracted from the users utterances and stored in a structured knowledge memory. The goal of each sub-dialogue is to retrieve a specific piece of information targeted by that particular coaching question.

Sub-dialogues aim to ensure local dialogue cohesion by contributing to a common ground, as described in Sect. 2.4.2. The corresponding coaching question provides the context for the subsequent set of turns. This fixed local context facilitates VPINO’s understanding of user utterances, information retrieval from these user utterances and planning suitable responses. Similar concepts have been presented for task-oriented conversation. [Grosz et al., 1977] suggest the general concept of sub-dialogues a subset of turns devoted to solve sub-tasks of the global problem.

SD	P	Text	DA
		...	
	1	Vpino: Ok, i think we are done with collecting options.	Statement
	2	Vpino: Do you have a favourite option yet?	W-Qu
	2	User: I am not really sure...	Pass
<i>Fav</i>	2	Vpino: Take your time and think about it.	Pump
	2	User: I guess the cheap one is my fav right now	Agree/Statement
	2	Vpino: Okay then.	Accept/Ack
	3	Vpino: "leasing" is your favourite.	Statement
	2	Vpino: Is that correct?	Y-N-Qu
<i>Cor</i>	2	User: yep	Agree
	2	Vpino: Okay,lets move on then.	Ack/Statement
		...	

**Figure 2.2:** Translated transcript of a real conversation from the decision coaching study. The example transcript includes two sub-dialogues (SD), the first on the identification of the favourite option of the user (*Fav*), and a short verification sub-dialogue (*Ver*). DA: Dialogue Act, P: Phase of the sub-dialogue (1) Grounding phase, (2) dialogue act sequence, (3) evaluation Phase.

### 2.5.1 Sub-dialogue Organisation

As already mentioned in the big picture in Sect. 2.1, a sub-dialogue is organised in three phases: (1) A leading grounding/presentation phase, (2) the dialogue act sequence model, and (3) the closing acceptance/evaluation phase. A real world example from our user study, translated from German, is shown in Fig. 2.2. The original German version of the conversation can be found in Fig. 2.3

The example transcript taken from a rational decision making session with VPINO covers two sub-dialogues. The first (*Fav*) aims to retrieve the users favourite option, the second (*Cor*) intends to verify the information extracted with in the first sub-dialogue. A schematic overview on the organization of a sub-dialogue in three phases is presented in Fig. 2.5. In the following, we provide a more detailed description on each of the three phases, following the example text in Fig. 2.2.

### 2.5.2 Grounding/Presentation Phase

The first phase is devoted to VPINO is setting up the context for the following turns and thereby contributing to common ground. Basically, VPINO posts a Statement on the actual state of the conversation or additional information the helps the user understand and answer the following question. Together with the original coaching question for a sub-dialogue, this first phase corresponds to the presentation phase in grounding theory.

SD	P	Text	DA
		...	
	1	Vpino: so dann sind wir fertig...	Statement
	2	Vpino: Welche der Optionen wäre denn ihr Favorit?	W-Qu
	2	User: hm ich bin nicht sicher...	Pass
<i>Fav</i>	2	Vpino: Lassen Sie sich Zeit und denken Sie ruhig nach!	Pump
	2	User: ja eigentlich ist die billige option die bessere	Agree/Statement
	2	Vpino: Okay	Accept/Ack
	3	Vpino: "eine auto leasen" ist also ihr Favorit.	Statement
	2	Vpino: Ist das Korrekt?	Y-N-Qu
<i>Cor</i>	2	User: jo	Agree
	2	Vpino: okay, dann mal weiter.	Ack/Statement
		...	

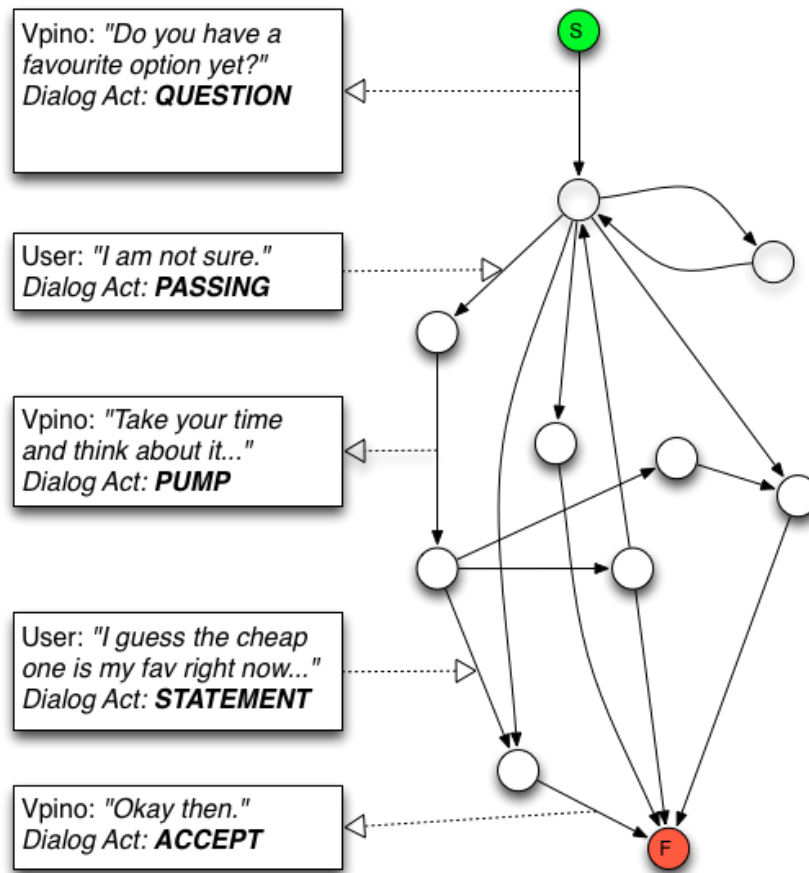
**Figure 2.3:** Original German transcript of the conversation for the example from Fig. 2.2. SD: Sub-Dialogue, P: Phase, DA: Dialogue Act.

In the example in Fig. 2.2, VPINO starts with a summary of the current dialogue state: The utterance “Ok, I think we are done with collecting options” informs the user that the previous sub-dialogue was successfully finished and the system wants to move on. The user is informed that the previous task of collecting options has finished. However, the pre-sub-dialogue grounding step is not mandatory. For example, the second sub-dialogue in Fig. 2.2 skips this phase, because there is no change in the context.

### 2.5.3 Dialogue Acts Sequence Phase

The actual conversation with the user is based on a *dialogue act sequence model (DASM)*. VPINO has a model for all expected dialogue act sequences that lead to an acceptable conversation state. The purpose of this model is to determine a valid response to a user message based on the current context of the conversation. The DASM is an extension of the adjacency pairs model on the set of dialogue acts as described in Sect. 2.2.3. For each sub-dialogue, it includes all expected sequences of valid turn-pairs of dialogue acts (as displayed in Fig. 2.2) to answer the corresponding coaching question. The DASM consists of a set of states and a set of transitions, i.e. the possible dialogue acts by both the user and VPINO at each state. Each sub-dialogue has a denoted initial and a final state. The initial state has a single outgoing arc, corresponding to the dialogue act of the coaching question. Paths through this graph represent valid dialogue act sequences by alternating authors. Acts by VPINO correspond to a set of preformulated text. This set consists of equivalent alternative formulations with the exact same dialogue acts, semantic meaning, and characteristics with respect to the conversation maxims. The actual text for the output





**Figure 2.4:** Example dialogue act sequence model (DASM) for the particular sub-dialogue on identifying a user's favourite option in a decision coaching session. Vpino states the coaching **Question**, the user is **Passing**. Vpino chooses the predefined possible dialogue act, pumping for information (**Pump**). The user finally answers with a **Statement**, which is accepted by VPINO (**Acknowledge**).

is selected randomly for output. The final state is reached when the user answered the question sufficiently, at least within terms of producing a valid dialogue act sequence.

Figure 2.4 shows a graphic representation of the DASM for the sub-dialogue used in the example from Fig. 2.2. After VPINO asked its initial question (i.e. ‘Do you have a favourite option?’) of the sub-dialogue, it is expecting one of the four possible dialogue acts (passing, agreement, disagreement, statement) as valid responses to that question.

In our example, the user’s response was classified as a *passing*. At that point of the sub-dialogue, VPINO assumes that a passing is not an acceptable final answer to close the sub-dialogue. The DASM defines a single valid response on a user passing, namely pumping the user for further information (*pump*). After the sequence question  $\rightarrow$  passing  $\rightarrow$  pump, VPINO is expecting either an Agreement, Disagreement or Statement by the user. In the example, the user responded with a formal acceptable response, namely the statement “I guess the cheap one is my favourite right now”. Note that the semantic content of that statement is not yet processed by the system. Nevertheless, VPINO displays that the message has been received and understood. Evaluation of the information of the dialogue act sequence follows in the dedicated evaluation phase. The final act leading to the final state is always an accept or acknowledgement by VPINO. For reasons of grounding (See Sect. 2.4), a DASM is always finished by an acknowledge/acceptance act by VPINO. Studies have shown that humans become confused when the system does not give acknowledge. The lack of acknowledgement in chat conversation can cause errors [Yankelovich et al., 1995].

#### 2.5.4 Evaluation Phase

In the final phase of the sub-dialogue, VPINO extracts and evaluates the requested piece of information. Therefore, the user utterances collected in the dialogue act sequence phase are parsed for information, as targeted by the initial question of the sub-dialogue. In case of successful evaluation, the SKM is updated with the new gained information. For reasons of grounding, the sub-dialogue can optionally show understanding in terms of displaying the information gathered.

For simple Y/N-questions, the target information is extracted by simply checking for an agreement or disagreement act by the user. For our example from Fig. 2.2, the targeted information was a semantic reference to a specific option, which the user probably described earlier in the conversation. In this particular case, the statement “I guess the cheap one is my favourite right now” is parsed and the reference to an option description, stored in the SKM, could be resolved. Finally, the SKM is updated with a reference to the option with the description “leasing a car”. In the example, VPINO now contributes to common ground by showing that the user utterance was understood, i.e. by repeating option description given by the user earlier. The task of semantic reference resolution will be subject of Sect. 5.6.

### 2.5.5 Problem Handling

Conversation with a human does not always go as planned by VPINO. In case of a misunderstanding or a failure of evaluating the extracted information, (e.g. the reference to the user's favourite option could not be resolved), VPINO is forced to immediately start a verification or correction sub-dialogue: A verification sub-dialogue is basically a precisely formulated yes/no-question, that asks the user to verify that a piece of information previously stored in the SKM is correct. A correction sub-dialogue tries to reformulate the original question and politely asks the user to repeat his answer in a more simple way. For the example above, VPINO then would initiate a follow up sub-dialogue that asks for a particular name or number of an option.

Unexpected or challenging user behaviour can result in misclassification of dialogue acts. The DASM only includes a set of dialogue acts by the user, which the author of the sub-dialogue expected to occur after a certain state. In case of an unexpected dialogue act performed by the user, VPINO is unable to interpret the intention of the user. In order to recover from unexpected (or potentially misclassified) user responses, VPINO enters a dedicated fall-back state. From this fall-back state, VPINO tries to "repair" the dialogue by pushing the user to answer the question properly. For example, VPINO reformulates the initial question and friendly asks the user give an answer.

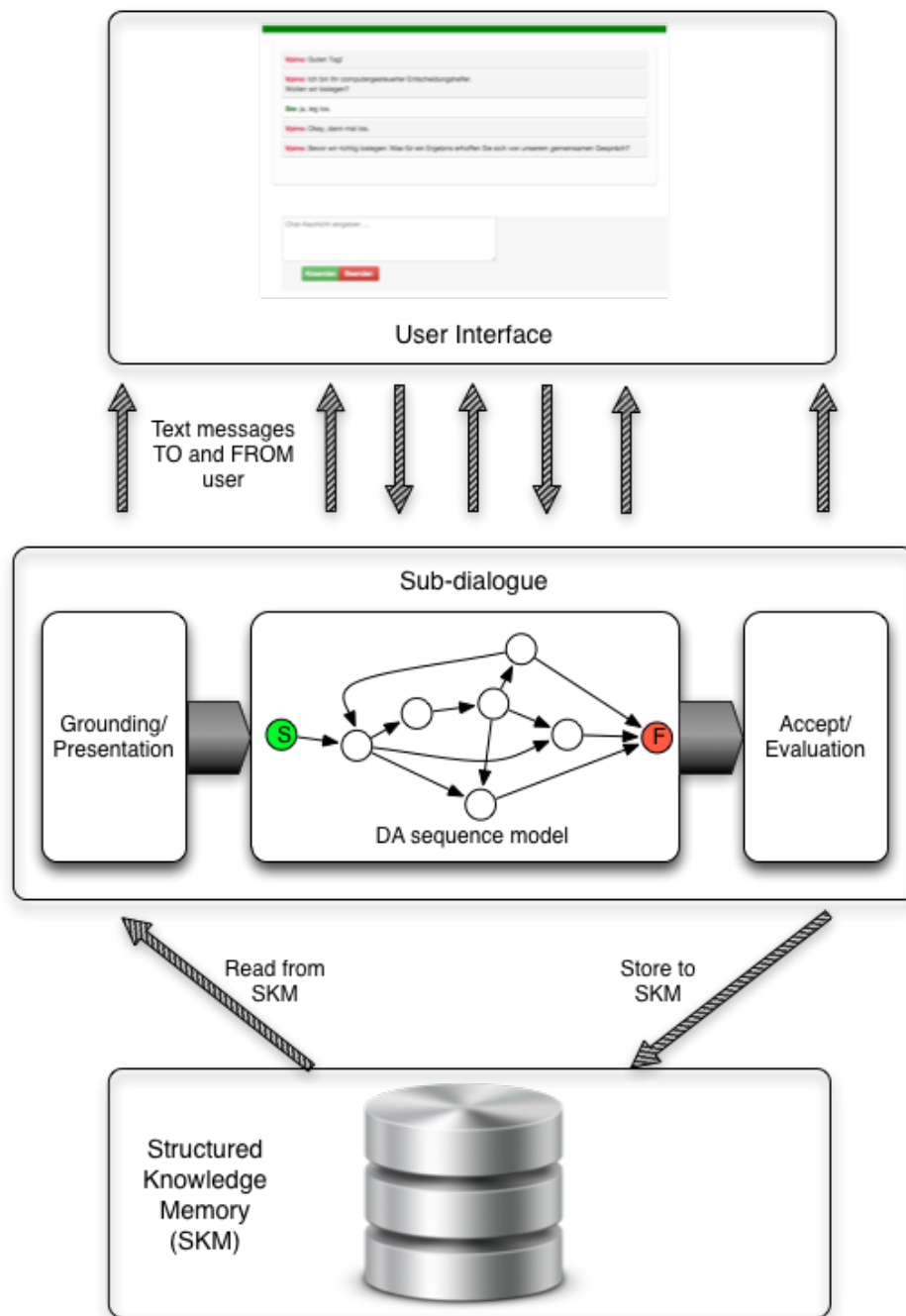
## 2.6 Communication and User Interface

While the previous sections focused on the conceptual model of a coaching dialogue and response generation by the system, this section will provide more detailed information about the components managing the actual communication with the user.

### 2.6.1 Timing and Turn-Taking

[Sacks et al., 1974] describe turn-taking as "A process by which interactants allocate the right or obligation to participate in an interactional activity." More precisely, turn-taking describes a local mechanism that determines the right to speak at *transition relevant places* in the conversation. These transition relevant places are often displayed by visual, prosodic or semantic hints in an utterance. For a text-based dialogue system, visual and prosodic features are not available. In contrast to spoken conversation, it is not even clear whether the communication partner is actually typing a message or rather waiting for a response.

On the other hand, in computer mediated text communication, such as chat, messages are transferred as a whole, in contrast to spoken systems, where barge-in and speech overlap is a problem for turn-taking mechanisms. However, latency in data transmission may lead to overlap for text based systems. To overcome these obstacles, a suitable turn-taking mechanism is required.



**Figure 2.5:** Schematic overview of a sub-dialogue. Processing starts with VPINO reading information from the SKM and presenting a grounding information to the user, immediately followed by a concrete coaching question/request. After reaching a final state in the DASM, VPINO extracts and evaluates the information from the preceding turns, updates the SKM and sends a grounding message to display “understanding”.

A common approach for conversational agents is the *request-response* ( or *stimulus-response*) approach. In general, a requestor (the user) sends a request message to a replier (the system) that ultimately returns a message in response. The protocol is a simple and frequently applied approach for communication management in dialogue systems. However, it is a simplification human of communication behaviour and therefore will not lead to human like conversation. On the other hand, it avoids problems arising from timing, interpretation of multiple messages or overlap.

### 2.6.2 Mixed-Initiative Communication

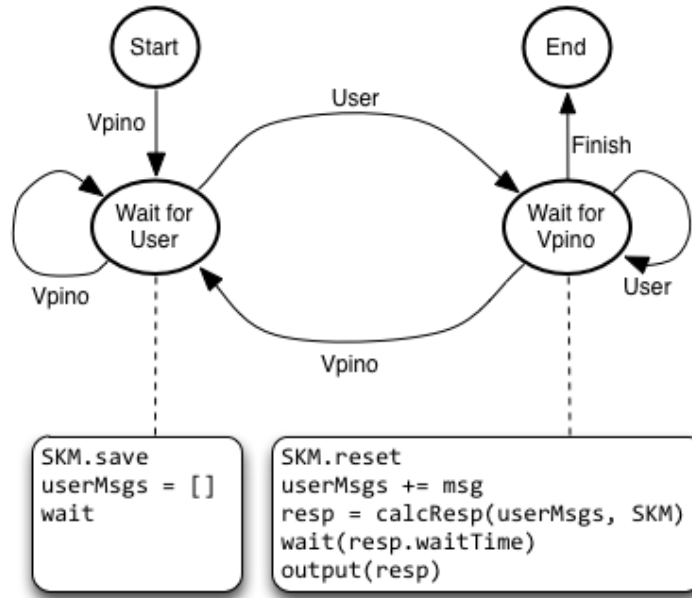
In human conversation, participants are free to contribute at any time they like. Moreover, human turn-taking behaviour “(...) is driven by social conventions and the interpretation of a large number of communication signals” [Duncan, 1972, p. 283]. There are no restrictions with respect to timing, order, or other strict turn taking rules. VPINO tries to imitate human chat behaviour. Therefore, we have implemented an asynchronous, mixed-initiative conversation protocol. Mixed initiative is a flexible interaction strategy in which “(...) both conversation partners contribute what is best suited at the most appropriate time”[Allen et al., 1999].

A main principle of VPINO is to keep control over the direction of the conversation at all times, by pushing the client in a passive role, such as as the client will not notice. VPINO keeps control over the direction of the conversation with by directing the conversation with pro-active behaviour and precisely formulated Socratic questions.

In contrast to its active role in directing the topics of the conversation, VPINO takes a passive role in turn-taking. Instead of instantly replying, VPINO waits for a short period of time and gives the user the opportunity to send additional messages in the meantime. A new message by the user forces VPINO to revise its pending answer - like a human that is interrupted while formulating a response. In contrast to the strict alternating order in the request-response approach, asynchronous communication allows the user to send multiple messages before the system responds to the user.

### 2.6.3 VPINO Turn-Taking Protocol

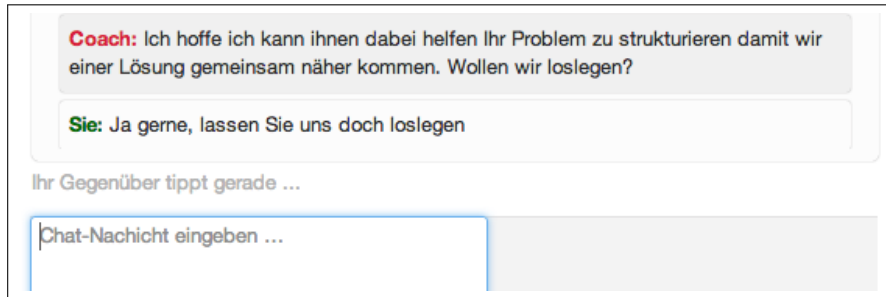
The turn-taking protocol intends to allow human-like mixed initiative communication, yet avoid misunderstandings that could result into loosing control over the conversation. Fig. 2.6 illustrates the state chart of the communication protocol. The protocol consists of a set of states, *Start*, *End*, *Wait for user* and *Wait for VPINO*. The set of transitions includes the messages sent by the *user* and VPINO, as well as a dedicated *finish* message from VPINO to finish the conversation. The pseudo-code in Fig. 2.6 describes a simplified version of the processes triggered by transitions from one state to another (or the same) one. Once a new message is received, the currently running process is interrupted. The process of the actual response calculation, i.e. the function *calcResp(messages,SKM)* in the pseudocode, is handled



**Figure 2.6:** A state chart for VPINO’s turn-taking and user communication protocol. Transitions are messages by the user or VPINO. Pseudo-code for transitions entering a state.

by the currently active sub-dialogue. The following list describes the states, their intended meaning and the particular process triggered on entrance.

- **Start** The initial starting point for conversation with VPINO. On startup, VPINO takes the initiative by immediately sending a greeting message (leading to the *Wait for User* state).
- **Wait for VPINO:** The state is entered after the user has sent a message. In case VPINO has already calculated a response to a message received earlier, but has not yet sent that response to the user, that response and the changes made on the SKM are obsolete and have to be re-thought. Conclusively, all temporary changes made on the SKM since VPINO’s last message are reset to the last state shared with the user. VPINO calculates a new response to all accumulated messages since the last message sent to the user. The resulting changes on the SKM, like the state of the sub-dialogue, are temporarily stored. The SKM will not be finally updated, until the calculated response is sent to the user and state *Wait for user* is entered. Before finally sending the message, the system waits for short time to give the user the chance to send a new message in the meantime. If no new user input arrived, the response message is finally sent to the user. Otherwise, the whole process is restarted.



**Figure 2.7:** User interface with text indicator for the virtual state of VPINO typing a response. “*Ihr Gegenüber tippt gerade...*” (in English: “*Your counterpart is currently typing...*”).

- **Wait for user:** VPINO has sent a message to the user. Once a message by VPINO is sent to (and seen by) the user, it cannot be revised. Therefore, the temporary state of SKM is finally saved. VPINO is now waiting for new user input.
- **Finish** The conversation is ended by a closing by VPINO.

#### 2.6.4 Waiting Time/Presentation Delay

VPINO waits for a short period of time and waits for further user input to simulate thinking or typing a message. The waiting time is calculated by a linear function based on the *word-count* ( $wc$ ) of the message. A minimum waiting time is set to simulate thinking up and typing a response. This ensures to give the user the opportunity to send further messages, even in case of very short responses by VPINO such as acknowledgements.

The maximum waiting time is set as a trade-off between human behaviour with possible unnecessary long waiting times on the one hand and usability on the other hand.

Throughout our studies we experimented with different timing settings for delayed responses. Our experience has shown that 1 second for the minimum, 5 seconds for the maximum with 0.3 seconds per word are well accepted by the users, resulting in the calculation function  $\min(\max((0.3\text{sec} * wc), 1\text{sec}), 5\text{sec})$ .

#### 2.6.5 Visual Delay Hint

Delayed system responses are intended to produce a more human-like feeling while talking to VPINO. However, many users expect a computer systems to respond instantly. As a result, delayed presentation can potentially lead to misunderstandings. Therefore, VPINO’s user interface was equipped with an additional visual hint. A short indicator text on the web interface, located between the user input field and the conversation history, as displayed in Fig. 2.7, visualises that VPINO is virtually typing a response.

## CHAPTER 2. BACKGROUND AND CONCEPTS BEHIND VPINO

The visual hint itself is also presented with a short time delay, since the presentation of the hint could possibly inhibit the user from typing additional information. Furthermore, as soon as the user has also started typing a new message, the visual hint is hidden. We do so in order to prevent misunderstandings.



## CHAPTER 3

# Systemic Coaching: Pre-study

---

The VPINO dialogue system was developed iteratively. In each iteration, we consecutively performed user tests and improved VPINO based on the experience with human users. Therefore, we conducted a first user study with a baseline prototype of our dialogue system in an early phase of development. The major goals of this study, apart from a proof of concept, were: (1) to get a first impression on how humans react to a natural language dialogue system in a coaching situation, (2) to discover potential differences to human-to-human conversation and finally, (3) testing and improving the coaching dialogue and dialogue act classifier. The results and insights from qualitative evaluation of the study had a large influence on the further development of the VPINO sub-dialogue model and dialogue act classifier.

This chapter is structured as follows: First, we will introduce “systemic coaching” in Sect. 3.1, followed by the adoption to VPINO in Sect. 3.2. In 3.3, we will describe the setting and results of our user study followed by the discussion and our conclusions in the last section.

### 3.1 Systemic Coaching

The systemic coaching approach is based on the sociological systems theory developed by Niklas Luhmann, a famous German sociologist [Luhmann, 1984]. From a systemic point of view, a human being is in constant interaction with the system in which it is currently located (professional, business environment, family, friends, hobbies, etc.). By changing the perspectives, problems can change or even dissolve, and there may be alternatives created. Thus, the problem-solving potential lies in the individual man himself.

Systemic coaching employs a question technique that is closely related to the technique of Socratic questioning. Sonja Radatz describes the technique as “Beratung ohne Ratschlag” [Radatz, 2003], which translates to “counseling without advice”. Like Socratic questions, the questions used in systemic coaching include rather open questions that aim to foster reflection: Circular questions ask for the clients presumptions attitudes and relations from others, in contrast to obvious questions about attitude, opinions and relationships of the client. Skaling questions use quantitative scales in order to reduce complexity in the conversation. Hypothetic questions such as what if questions allow the client to construct a different reality and allow to develop action plans how to either achieve or avoid this scenario. Steve de Shazer suggests “solution-oriented questions to define a desired future instead of thinking to much of the problem” [De Shazer and Dolan, 2012].

### 3.2 Vpino Dialogue Plan for Systemic Coaching

This first prototype did not yet include the sub-dialogue model introduced in Sect. 2.5. However, a baseline version of the rule based dialogue act classifier was available. Also, communication with VPINO followed a basic request-response protocol. We use a dialogue plan based on the following phases employing well established systemic coaching questions as presented in by [Radatz, 2003] and [De Shazer and Dolan, 2012]. It consists of the following phases:

- Framing the subject of the conversation
- Problem Exploration
- Changing the perspective
- Identify goals and desired future
- Plan next steps

In the first phase, we frame the conversation; i.e. ask for the clients expectations, explain our methodology and clarify what we will not do with the conversation, i.e. giving advice on their specific problem. The second phase is held rather short and is devoted to the

problem exploration. Thereafter, we change the perspective by asking how others would rate and handle their problem. We try to identify negative aspects and action patters that would make the problem even worse. Finally, we ask the client to define how a desired future would look like and ask her to develop a set of action strategies to achieve an optimal solution. The exact set of questions used in our study can be found in the Appendix.

## 3.3 User Study

We tested the general problem coach with a group of 32 students attending their last year in school (age between 16 and 18 years). The subject of the conversation is a general purpose problem coaching with a focus on career decisions. The participants were split into two groups: The first group (N = 22) held conversations with VPINO, the latter (N = 10) received a chat coaching by a human. Generally, the human coach used the same set of systemic questions and followed the same dialogue plan. Nonetheless, he was allowed to respond accordingly if required by the situation. From the user's perspective, the presentation of the chat-coach was identical, both groups used the same web-interface. However, due to the small number of participants and the early state of development a quantitative evaluation was not performed. Instead, we evaluated our system by reviewing the conversation transcripts.

### 3.3.1 Results

Five out of the 22 conversations with VPINO produced nearly perfect human like dialogues without any major problems. Apart from technical problems, for example misclassification of user responses, many ineffective/problematic conversations resulted from the clients behaviour: Some of the participants were not willing to accept the restrictions of our system. Some of the users confused systemic coaching with giving advice and started the dialogue with questions like "What should I do after finishing School?" instead of responding to VPINO's questions accordingly. Needless to say, VPINO neither intends nor is it capable to answer such a question. For example, a small number of participants intentionally challenged the systems natural language understanding capabilities by repeatedly asking questions, although constantly reminded by VPINO that it will not be answering those questions. However, participants that were willing to work with our system and truly interested in finding a solution for their problem did hold fruitful conversations with VPINO.

### 3.3.2 Naturalness of Conversations

Comparison between high quality VPINO conversations with human chat coaching and low quality VPINO conversations revealed interesting insights about the naturalness of

these conversations: Most of the high quality conversations with VPINO resulted from detailed and cognitive reflected answers by the clients. In contrast, participants that answered in a few words only did not produce natural conversations. After looking at the human conversations with rather short client responses, we found that these included a lot of social clues and acknowledgements. The human coaches tend to signal understanding what seemed to help a lot in keeping the participants focus on the conversation. For human conversations with long and reflected user responses, it was sufficient to stick with the systemic questions.

### 3.4 Discussion and Conclusions

Over all, we were satisfied to successfully help motivated users with their problem. Even with a rather simple, straight forward implementation of a dialogue act based coaching conversation, VPINO was able to hold an effective conversation with some user.

Nonetheless, there were still some issues that future work has to address. Systemic coaching questions have a rather open nature, which makes it hard to follow the user's thoughts and direct a structured conversation. Consequently, a general systemic coaching approach is not an optimal scenario for computer based coaching, especially when using shallow natural language understanding on the level of dialogue acts. Precise, targeted formulations are required in order to avoid misunderstandings. Efficient fall-back mechanisms in case of dialogue act misclassification are required. Complex counter questions that the system does not understand also need to be handled properly with repair strategies. From Sect. 3.3.2, we conclude that there is a strong need for social cues and grounding strategies to provide more natural conversation that lead to effective problem solving. Also, attitude and personality matters: The quality and effectiveness seemed to strongly depend on user cooperation and have to be examined more detailed in further studies.

Most important, the pre-study provided us with a large number of response samples from human users. These sample text could be successfully used to develop the dialogue act sequence models, improve the dialogue act classification rules in the rational decision coaching study (See Sect. 5.4).

## CHAPTER 4

# Training Transfer Coaching: User Study

---

In this Chapter, we present the results of a user study for the scenario of VPINO as a training transfer coach. Participants received an online communication training followed by a daily training transfer coaching by VPINO for one week. The aim of this study was to examine the usefulness of computer-based transfer coaching as a training transfer strategy. For training of communication skills, we used the *Simulation-based Training on Real-life Incidents* (STORI) tool that was developed at the department of psychology at TU Darmstadt. In a longitudinal experiment, we tested the transfer coaching by VPINO against a conventional strategy (daily filling out a daily online survey) with regard to their effectiveness on communication skills. Based on the insights from our first experiments (See Sect. 3), we also examined whether the user's personality trait openness is a key factor for the effectiveness of a computer based coach. Parts of the user work presented in this chapter have been published in [Mäurer et al., 2013]

This chapter is structured as follows: The chapter starts with background information on training transfer in Sect. 4.1, then continues with the influence of personality on coaching success in Sect. 4.2. Sect. 4.3 presents a sketch of the online communication training tool STORI, followed by the basic concepts of transfer coaching and their adoption to a dialogue plan for VPINO in Sect. 4.4. The effectiveness of VPINO as a transfer coach was tested in a user study. Measures and results of this study will be presented in Sect. 4.6.

## 4.1 Background and Motivation

Training is successful, if training transfer was successful [Barnett and Ceci, 2002]. Training transfer is defined as “...the degree to which trainees effectively apply the knowledge, skills, and attitudes gained in a training context to the job. For transfer to have occurred, learned behaviour must be generalized to the job context and maintained over a period of time on the job.” [Baldwin and Ford, 1988, p. 63]. Bridging the gap between training and everyday work is especially important for soft skills (e.g. communications skills) as the training can only cover a small percentage of the situations in which social skills are required [Keith et al., 2010]. However, at work, costs of failure and the pressure to meet deadlines typically hinder employees from exploring new and alternative methods. Instead of further improving their skills, they tend to rely on existing and well-practiced methods [Ericsson et al., 1993, Haccoun, 1997]. Transfer interventions are effectively used to increase the motivation of learners to use their newly acquired skills in their daily routine. In particular, literature strongly supports the use of the goal setting strategy [Burke and Hutchins, 2007]. Comparative studies reveal that goal setting is superior to other post-training interventions in terms of increasing trainees’ transfer performance [Wexley and Baldwin, 1986]. As solution-based brief coaching [Berg and Szabo, 2005] is a particular form of goal setting, we suggest coaching to be a suitable alternative to conventional post training interventions (e.g. goal setting via free-text input in web-form). Decisive superiority of coaching in comparison to other transfer strategies may lie in its ability to enhance participants’ meta-cognitions [Grant, 2003]. Meta-cognitions capture the planning how to best achieve a specific goal, monitoring the progress and the evaluation of the used strategies [Schraw and Moshman, 1995]. First research attempts were able to show that managers who received personal coaching after a training intervention further increased their productivity during the coaching phase [Olivero et al., 1997]. However, personal coaching would be far too expensive in order to provide it to a larger number of employees.

Our computer-based transfer coaching with VPINO, on the contrary, could be an effective and economic alternative.

## 4.2 Does Personality Affect the Success of Coaching?

We expect personality to affect the success of computer-based transfer coaching. “Openness to new experience” is one of the “Big Five” personality factors. It encompasses intellectual curiosity, preference for variety and the willingness to explore new ways [Costa and McCrae, 1992]. Especially for complex and changing task conditions, openness has shown to significantly impact the effectiveness of training interventions [Herold et al., 2002].

Furthermore, openness has a significant positive influence on coaching success [Stewart et al., 2008] and certain components of user acceptance [Devaraj et al., 2008]. In contrast to conventional transfer strategies (e.g. multiple choice online form sheets), we expect user ac-

ceptance to be a key factor for the success of computer-based transfer coaching. Therefore, we argue that individuals with higher openness will benefit more from computer-based transfer coaching than individuals with low openness.

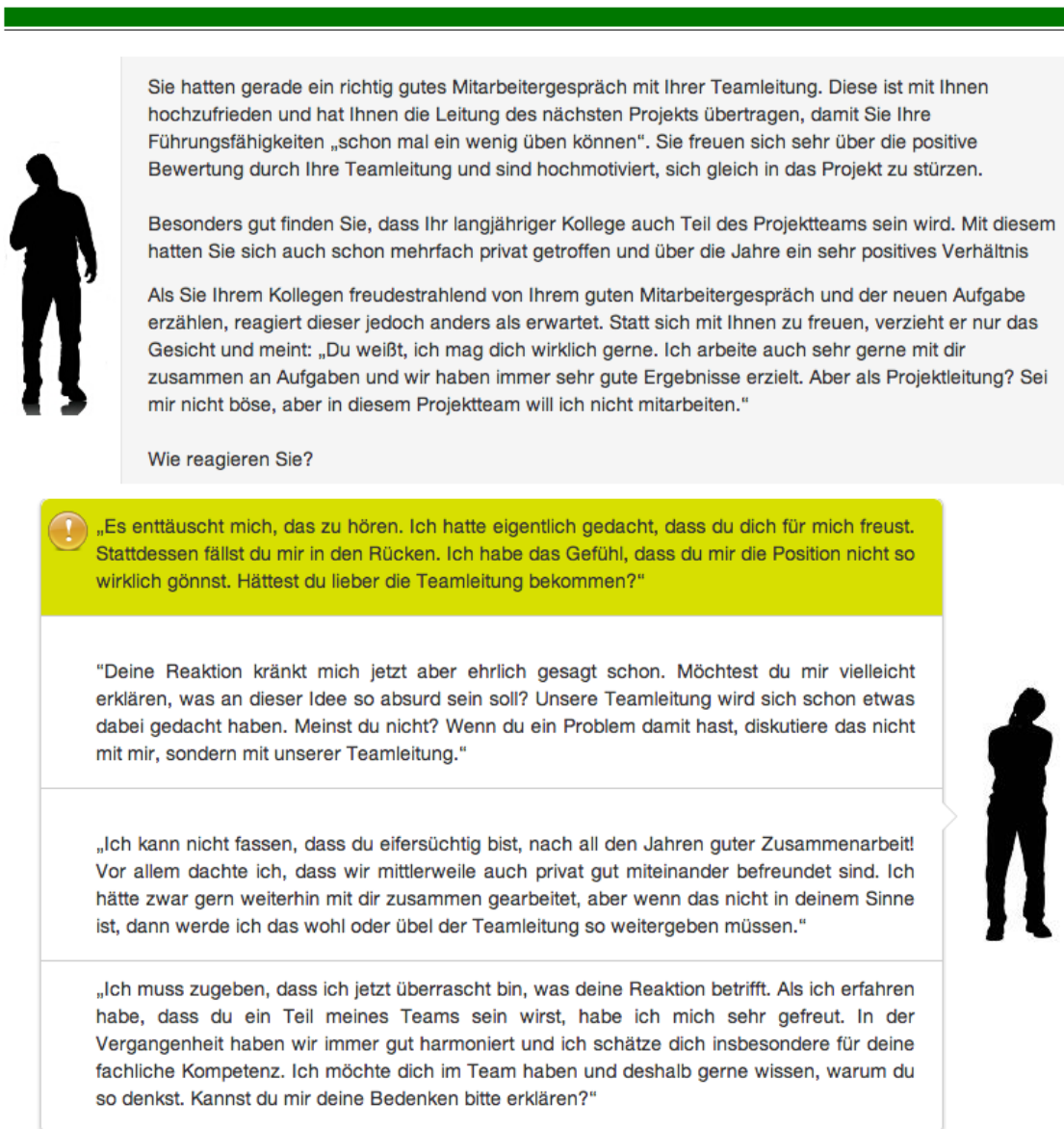
### 4.3 Simulation-based Training on Real-life Incidents

The communication skills training was provided with the *Simulation-based Training on Real-life Incidents (STORI)*. It was developed at the department of psychology at TU Darmstadt and is based on a collection of situational judgement tests (SJTs). SJTs can be successfully applied for communications skills training by simulating critical situations. The focus of the training is general procedural knowledge about effective and ineffective communication in critical professional situations.

The critical situations were adapted to short multi-option dialogues in order to simulate a conversation between the trainee and another person. The resulting SJTs are presented to the trainees as a text based dialogue with a list of preformulated response options. Fig. 4.1 shows an example of such a dialogue simulation. The trainees were also provided with feedback on their selected responses by coloured highlight of their choice.

The advantage of this simulation-based training exercises is the proximity to real-world scenarios. This realism significantly facilitates the transfer of learning into everyday work. In conventional training methods, training content and related exercises are processed in sequential turns. In contrast, SJT's require targeted integration of specific content learned. The trainee is forced to independently distinguish between important and unimportant information to make a decision while exercising. Additionally, various case studies can be linked to a story. Thereby, the consequences of the decision taken will also immediately be confirmed, which also greatly supports the learning progress. Simulation-based training is particularly suitable for situations without single or clear solutions, such as communication or negotiation training.

The collection of SJTs for *STORI* was developed using the *critical incident technique*: In a first step, case studies of critical communicative situations in work life were collected. In a second step, 30 cases that are considered as particularly critical were selected. For these 30 cases, a number of response options to the critical situation were formulated, each with a slightly varying "goodness" or degree of communication quality. These response options were evaluated by professionals to ensure the proximity to everyday work. At the same time, these professionals were instructed to formulate reactions to each particular response options again. This approach was repeated to construct SJTs for simulated conversations with five parts. (1) The initial presentation of the situation by *STORI*, (2) a user reaction, (3) a response from the virtual communication partner, (4) a second user reaction followed by (5) a final closing to end this short "story". Overall the development was accompanied and supported by the input of more than 150 professionals.



**Figure 4.1:** User interface of STORI communication training tool. In the displayed example situation, the user is confronted with a description of a critical communication incident (in German). A summary of the situation displayed translates as follows: The trainee is put into the new position of a team-leader. When happily telling his colleague about it, the colleague reacts unexpected sceptical. The colleague states that he does not want to work in the trainee’s team and is questioning the trainees competences for that job. The trainee chooses from four possible reactions that vary with respect to the displayed emotion, in this particular case disappointment, anger, insult and surprise, and how constructive the trainee resolves this critic situation. Immediate feedback is provided by coloured highlighting of the user selection.



## 4.4 Solution Based Brief Coaching

For the training transfer scenario, we employ a coaching approach called *Solution Based Brief Coaching* introduced by [Berg and Szabo, 2005]. It supports the client in defining specific goals and finding a suitable goal accomplishment strategy. Also, brief coaching aims to achieve these goals in a comparatively short space of time [Ives, 2008]. Normally, it focuses on a relatively defined issue or goal, in contrast to more general/global goals issued with systemic coaching approaches. Brief coaching thereby makes use of a number of methods that support this strategy, for example looking for resources rather than deficits, exploring possible and preferred future states and examining what is already contributing to that preferred future. Finally, brief coaching principles treat clients as experts in all aspects of their lives. Hence, it fits as a strategy for communication training transfer in accordance with VPINO's design principles. A coaching session in *brief coaching* encompasses three stage phases:

1. **Define a desired future:** Defining a specific goal (in life) the user wants to reach with the help of the coach.
2. **Changes in State:** Discuss recent past, look for indicators of changes in direction of desired state.
3. **Experimental Phase:** How to put into practice what has been discussed so far, agree on minor changes in everyday activity.

## 4.5 VPINO Dialogue Plan for Brief Coaching

The principles of brief coaching were adapted in the dialogue plan for training transfer with VPINO. For our training transfer scenario, the user has to hold a daily conversation with VPINO for the period of one work-week, which results in up to five conversations. After communication training with STORI on the weekend, a first initial conversation of the week was used to introduce VPINO to the trainees, whereas the following conversations particularly focused on evaluation of the trainees daily progress.

### 4.5.1 Initial Conversation

In the first coaching session of the week, VPINO is introduced to the trainees. Afterwards, the trainees are asked to set their goal for the whole week, i.e. which particular communications skills they want to improve. The trainees are also asked about how they estimate their opportunities/chances to apply the newly acquired skills to work. The basic idea of this first conversation is to sensitize the trainees on critical situations throughout the week and motivate them to actively apply the lessons learned from the training. The first conversation closes with setting a particular goal for the next workday.

### 4.5.2 Recurring Daily Conversations

The conversations on the following workdays focus on evaluation of the users weekly and daily goals. VPINO accesses the information stored in the SKM of the preceding conversations in order to reference the weekly/daily goals that were set in these previous sessions. The trainee is asked to explain whether or not she had the chance to work on her goals. If she was not, she is asked to name the reasons. The dialogue plan included questions on the trainees general progress achieving her goals. In case of little or no progress at all, the user is asked to identify possible obstacles. In order to motivate the trainees, one part of the dialogue is devoted to identify and point out positive experience with their work on communication skills. In the final part of the daily conversation, VPINO aims to foster the trainee's reflection about the effect of his work on his personal goals. To give you an idea how a training transfer session with VPINO looks like in practice, we present the translation of a workday session transcript from German in Fig. 4.2.

Note that the daily conversations are all based on the identical dialogue plan containing the same set of coaching questions. Nevertheless, the resulting conversations with the trainee differ significantly from day to day: The selection of questions by the dialogue manager depends on the goals set by the trainee, the progress made the previous day and finally their experience during that particular workday. Even if the same sequence of questions would be selected by the dialogue manager, the preformulated text used in VPINO's utterance varies.

## 4.6 User Study

### 4.6.1 Study Design

Our study was conducted as a longitudinal field experiment with three measurement points. Participants were 26 alumni of TU Darmstadt (50% female) who voluntarily signed up for an online communication training program. On average, participants were 36 years old ( $SD = 10.89$ ). They completed one to five communication trainings prior to this training in their professional lives.

Our study was conducted in two phases, as displayed in Fig. 4.3: An online survey on participants' Communication Skills (T1) was followed by the training phase. The training phase ended with a post-test on Communication Skills (T2). Immediately after the post-test, participants were randomly assigned to two different experimental conditions: Twelve participants were instructed to record their progress in an multiple choice online form sheet on a daily basis, over the course of one week. The remaining fourteen participants were instructed to use the computer-based transfer coaching also on a daily basis over the course of one week. The effects of the different transfer strategies were assessed in a follow-up test on participants' Communication Skills (T3) after the end of the transfer week.

**Coach:** Hello again. The last time we talked to each other you told me you couldn't work much on your goal. Did you come any closer in the meantime?

**Trainee:** Yes, whenever a mistake occurs, I try to stay calm and analyze the situation.

**Coach:** Ok. How would you rate yesterdays progress on your weekly goal? Did you fully reach your goal or only to some extent?

**Trainee:** Yesterday, one my supervisor talked to me about a mistake I made.

**Coach:** Go on.

**Trainee:** I remained calm and analyzed the situation. Based on my analysis I recognized that my mistake can be corrected.

**Coach:** And how would you rate your goal accomplishment in percent? (0 - 100)?

**Trainee:** I would say about 80%.

**Coach:** Congratulations, well done! Why do you think you haven't accomplished more than that?

**Trainee:** At first, it is always a shock and pretty difficult to stay calm. But when I take the time and analyze the situation, its getting better.

**Coach:** Is there anything that went particularly good, when you think about working on your goal yesterday?

**Trainee:** Yes, that i remained calm although I was stressed.

**Coach:** What exactly was positive about that situation?

**Trainee:** To stay calm and not get hectic

**Coach:** Ok good. Now that we have reflected what went well and not so well on your last workday is there something that you would like to change tomorrow?

**Trainee:** No, I think its okay. I'm starting to get a routine in such situations.

**Coach:** And what could help you reach your weekly goal completely?

**Trainee:** I guess thats not really possible in the near future.

**Coach:** Are you satisfied with your actual goal accomplishment?

**Trainee:** I think I am.

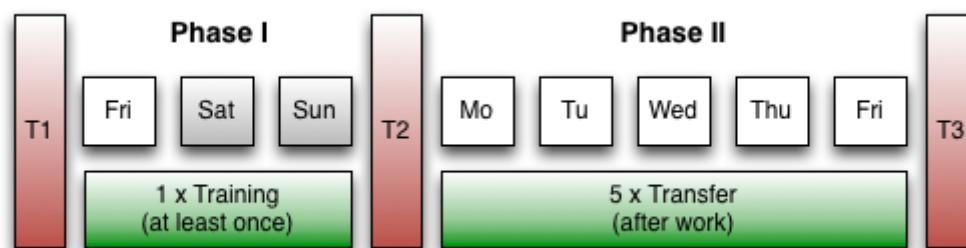
**Coach:** That sounds good. And how much would you say does the work on your goal affect your communication skills? (strong to not at all)

**Trainee:** I would say strong...

**Coach:** Alright. I think that's it for today. Goodbye!

**Trainee:** Bye

**Figure 4.2:** Transcript of a training transfer session taken from the user study, translated from German. The presented transcript was the third conversation in the transfer week. Note that multiple messages by VPINO and the user have been condensed to single messages due to space restrictions in this figure.



**Figure 4.3:** Diagram for the user study design for training transfer coaching. Measurement of communication skills in T1, followed by the communication training in phase 1. Assessment of communication skills in T2, followed by a week of transfer coaching, ending with a final assessment of communications skills in T3.

Instructions after the post-test (T2) in both experimental conditions (Computer Based Coaching and Online Form Sheet) were **identical** in the following aspects: On the first day (after the post-test at T2) participants were asked to define a specific goal they want to accomplish in the transfer phase. This goal should refer the improvement of communication skills. Positive examples for goals were provided. Participants were also asked to rate the feasibility, and to name potential promoters and inhibitors to accomplish their goal. On the following four days participants were instructed to rate goal accomplishment and the (positive or negative) effect of this goal accomplishment on their communication skills. Furthermore we asked participants to name the specific promoters and inhibitors of goal accomplishment they faced during that day.

Instructions after the post test (T2) for both experimental conditions (Computer Based Coaching and Online Form Sheet) were **different** in the following aspect only: Whereas the Online Form Sheet presented the instructions in a static form, the Computer Based Coaching presented the questions adaptively in the form of a coaching conversation with VPINO, as described in Section 4.5.

#### 4.6.2 Measures

Communication skills were assessed in a test at all three measurement points: in a on-line survey prior to the training phase (T1), in a post-test after the communication training phase, prior to the experimental manipulation of the Transfer Strategies (T2), and in a follow-up survey after the experimental manipulation of the Transfer Strategies (T3).

The test consisted of three critical situations that were presented to the participants (i.e. 9 critical situations in total). Within 15 minutes, participants had to generate as many useful and original responses to the given situations as possible. Two independent experts rated the quality of the different responses on two dimensions (usefulness and originality)

on an anchored 7-point Likert scale. Multiple responses of an individual participant were averaged per situation and dimension. A single Communication Skills Index was formed by multiplying scores on these two dimensions [Zhou and Oldham, 2001]. A global Communication Skills Index per measurement point was aggregated across the three test situations. This elaborate procedure resulted in a good agreement between the ratings of the two independent experts (ICC .70 to .84).

Openness was measured by two items derived from [Rammstedt and John, 2007] Big Five Inventory-10. Participants rated themselves on both items ('I see myself as someone who has as few artistic interests.' and 'I see myself as someone who has an active imagination.') on a five-point Likert scale (1= disagree strongly to 5 = agree strongly). Both items were later aggregated to a global Openness score.

#### 4.6.3 Control Variables

In our analysis, we wanted to see the “pure” effect of our Transfer Strategy (Computer Based Coaching and Online Form Sheet) without the distortion of other influential factors. Therefore we controlled for several variables in our analysis that we expected to also have an influence on Communication skills at T3 apart from our Transfer Strategy:

First, we expected our participants to differ in their Communications Skills prior to the training (at T1) and prior to the experimental manipulation of the Transfer Strategy (T2). In order to eliminate both the influence of prior Communications Skills and the effects of the training itself, we included Communication Skills at T1 and T2 as control variables into our analysis.

Second, we provided participants with access to the training chapters also after they had completed the post-test at T2. As further repetition of the training chapters may also cause a further improvement of participants' Communication Skills, we recorded the Login Frequency after the post-test and controlled for its influence in our analysis.

Third, we expected the participants' individual motivation to have an influence on the effectiveness of the training and possibly interfere with the effects of the different Transfer Strategies. Therefore, we assessed the participants' initial motivation to sign up for our communication training on 15 items (e.g. “My main motivation to participate in the training is because I want to improve my social skills”). These items covered five dimensions of motivators from technical aspects to career advancement. Participants rated their motivation on a five-point Likert scale (1 = disagree strongly to 5 = agree strongly). All items were later aggregated to a global *Motivation score*.

## 4.7 Results

### 4.7.1 Descriptives

Our analysis encompassed one dependent variable (Communication Skills at T3), two independent variables (Openness and Transfer Strategy) and four Control Variables (Communication Skills at T1, Communication Skills at T2, Login Frequency and Motivation). Means, standard deviations, and inter-correlations among all variables are presented in Table 4.1. Communication Skills are significantly correlated ( $p < .05$ ) across the three measurement points. Furthermore, the Transfer Strategy and Login Frequency were significantly correlated ( $p < .10$ ). As Computer Based Coaching was contrast coded with +1 and the Online Form Sheet with -1, this positive correlation indicates that participants in the *Computer Based Coaching* condition had more logins after the post-test at T2 than participants in the Online Form Sheet condition. All other variables did not differ significantly between the two experimental conditions.

### 4.7.2 Quantitative Results

	Total M (SD)	form M (SD)	coach M (SD)	1	2	3	4	5	6
Transfer Strategy	-0.08 (1.02)	1.00 (0.00)	-1.00 (0.00)	-					
Commun. Skills (T1)	19.86 (4.83)	19.94 (5.07)	19.84 (4.81)	-.01	-				
Commun. Skills (T2)	20.68 (6.09)	20.28 (5.42)	21.02 (6.79)	.06	.40*	-			
Commun. Skills (T3)	22.57 (4.32)	22.43 (3.34)	22.68 (5.14)	.03	.44*	.49*	-		
Openness	2.54 (1.07)	2.42 (1.38)	2.64 (0.74)	.11	-.07	.10	.03	-	
Motivation	2.20 (0.51)	2.26 (0.60)	2.15 (0.43)	-.10	.33	.31	.22	.18	-
Login Frequency	4.23 (4.62)	2.50 (1.73)	5.71 (5.77)	.35 <sup>T</sup>	-.17	-.08	-.28	.19	-.06

**Table 4.1:** Means, Standard Deviations and Pearson Correlations. Note: *Computer Based Coaching vs. Online Form Sheet* was contrast coded: Computer Based Coaching (coach) = 1; Online Form Sheet (form) = -1. <sup>T</sup> $p < .10$ , \* $p < .05$ .

We assumed that participants who are more open to new experience will benefit more from Computer Based Coaching than participants who are less open to new experience.

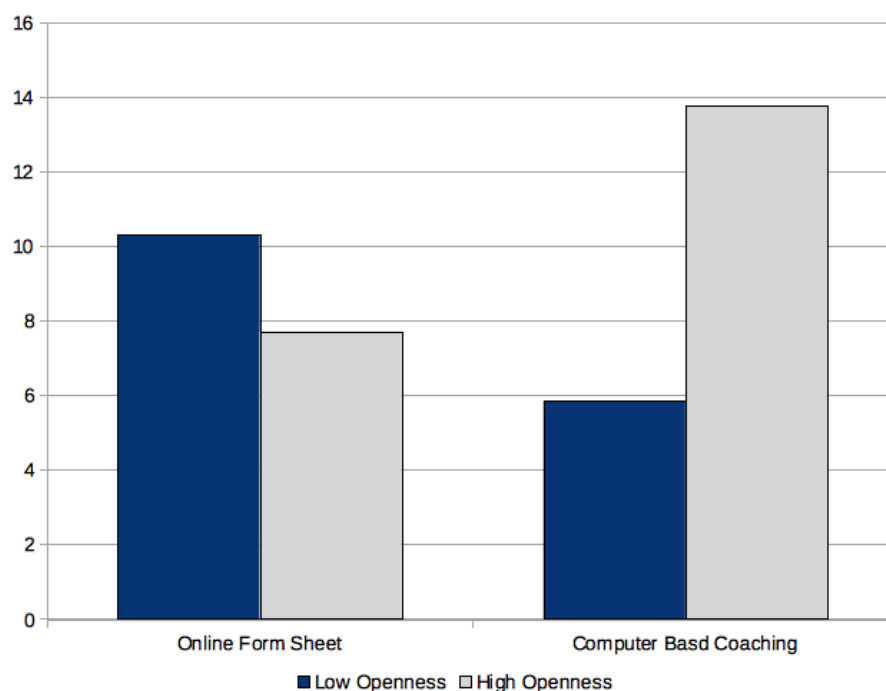
	Step 1	Step 2	Step 3
Change in variance accounted for Communication Skills at T3 ( $\Delta R^2$ )	.36*	.01	.14*
Communication Skills (T1)	0.26	0.26	0.12
Communication Skills (T2)	0.37 <sup>T</sup>	0.36 <sup>T</sup>	0.50*
Login Frequency	-0.21	-0.25	-0.33 <sup>T</sup>
Motivation	0.01	0.01	0.28
Transfer Strat.	—	0.09	0.08
Openness	—	0.05	0.26
Transfer Strat. x Openness	—	—	0.55*

**Table 4.2:** Test of Moderation Transfer Strategy x Openness on Communication Skills. Note: Values are standardized regression coefficients. All terms were centered prior to analysis. <sup>T</sup> :  $p < .10$  \*:  $p < .05$ .

We tested our assumption using hierarchical regression analysis. In the first step, we entered the control variables (Communication Skills at T1, Communications Skills at T2, Login Frequency and Motivation). In the second step, we entered the moderator variable (Openness) and Transfer Strategy (*Computer Based Coaching vs. Online Form Sheet*). The interaction term between Transfer Strategy and Openness was entered in the third step [Aiken and West, 1991]. To reduce multicollinearity, all variables were centred at their respective means.

Table 4.2 reports the test of our assumption: The Control variables entered in step 1 of the hierarchical regression analysis already accounted for 36% of the variance in the Communications Skills at T3. In particular, Communications Skills at T2 significantly predict Communications Skills at T3 ( $\beta = .37, p < .10$ ). When we entered the Transfer Strategy (Online Form Sheet vs. Computer Based Coaching) and Openness in step 2 no additional variance in the Communications Skills at T3 was explained. However, when we entered the interaction between Transfer Strategy and Openness in Step 3 additional 14% of the variance in the Communications Skills at T3 could be explained. In sum, a total of 50% of the variance in the Communications Skills at T3 can be explained by using this set of variables. Among the variables three significant predictors were identified: The interaction between Transfer Strategy and Openness significantly predicted Communication Skills at T3 together with the Control variables Communication Skills at T2 ( $\beta = .50, p < .05$ ) and Login Frequency ( $\beta = -.33, p < .10$ ).

The plot of the relationship between Transfer Strategy and Openness is presented in Figure 4.4 and supported our hypothesis: Participants who are more open to new experience benefit more from the Computer Based Coaching-condition than participants who are less open to new experience. The simple slope analysis revealed this difference to be



**Figure 4.4:** Gain in Communication Skills at T3 depending on Openness and Transfer Strategy after controlling for Communication Skills at T1, Communication Skills at T2, Login Frequency and Motivation.

significant ( $p < .05$ ). In the Online Form Sheet-condition, the effect seemed to be reversed. However, the simple slope analysis revealed this difference not to be significant (*ns.*).

Apart from its effect on communication skills, we examined the quality of the coaching dialogues with regard to their naturalness. Most of the conversations held with VPINO were rated quite natural by the users.

### 4.7.3 Qualitative Results

A review of the conversation transcripts showed that in a professional setting, the users seemed to be more indulgent compared to our pre-study. The participants were more likely to tolerate minor misunderstandings with VPINO and thus made it easier for VPINO to recover and continue the conversation.

With some of the participants, the recurring conversations based on a similar set of questions resulted in a routine when speaking to VPINO. The length of these users' responses often reduced to concise utterances towards the end of the week. This behaviour was observed more often with participants that did not manage to work on their daily or weekly goals very often. Even so, in the particular case of no progress on the trainee's



communication skills, the conversations with VPINO are rather short since there was not much to talk and reflect about anyway. Also, routinised user behaviour also arises with filling out the recurring online form-sheet every day.

## 4.8 Discussion and Conclusions

With our user study, we investigated the effectiveness of the VPINO chat coach in the professional setting of training transfer coaching. Our results suggest that computer-based transfer coaching with VPINO can effectively help participants to further increase their communication skills after a training intervention. Also, the higher login rate for VPINO suggests that a chat conversation with VPINO was more motivating than online form sheets.

However, personality has an impact on the success of using VPINO as a transfer coach. Participants largely differ with regard to their openness: Participants with high levels of openness benefit more from computer-based transfer coaching than participants with low levels of openness. Nonetheless, openness to new experience does not seem to influence the effectiveness of the conventional multiple choice online form sheet. This implies that computer-based transfer coaching is probably not suitable for everybody. Users that remain sceptical towards computer-based transfer coaching will not benefit from using this technology. Future work needs to further explore the influence of personality variables on the effectiveness of conversational agents. A failure to consider personality in future research may even bias the results of similar studies. To conclude, a natural language dialogue system such as VPINO can be successfully applied in a real world coaching scenario.

## CHAPTER 4. TRAINING TRANSFER COACHING: USER STUDY

## CHAPTER 5

# Rational Decision Coaching with VPINO

---

In this chapter, we show that rational decision-making methods can be successfully implemented as a text-based natural language dialogue system. For employ a goal oriented decision coaching approach for VPINO. It is based on the method of rational decision-making as first introduced by Benjamin Franklin [Sparks, 1844], also called *Pros and Cons*. We describe the adoption of Franklins’s method for VPINO and the representation of the user’s decision problem in the SKM. Furthermore, this chapter provides implementation details on specific tasks related to computer based decision coaching.

**This chapter is structured as follows:** Background information on rational decision-making is provided in Sect. 5.1. Benjamin Franklin’s approach on decision-making is presented in Sect. 5.2. Sect. 5.4 describes the adoption of Franklin’s decision-making approach for a coaching conversation with VPINO. Details on VPINO’s method for pairwise option comparison are presented in Sect. 5.5. VPINO’s method for resolving option and argument references is described in Sect. 5.6.

## 5.1 Background and Motivation

### 5.1.1 Rational Decision Making

Decision-making is commonly defined as a cognitive process with the goal to select a final choice from a set of options. Decision-making is an elementary and recurring task for humans. It involves handling information, uncertainty and resources and is also influenced by several psychological factors such as habits or personality. In this work, the term decision-making is used in terms of rational *choice* between a set of alternatives from an economic perspective (in contrast to the neurologic processes responsible for human behaviour). Oskar Morgenstern and John von Neumann, founders of game theory, defined rationality as a process of following normative decision rules, for example choosing the option with the most promising subjective utility, or showing consistent preferences [Von Neumann and Morgenstern, 1944]. However, rational decisions can vary on a broad range from personal decisions up to management decisions with a potentially high impact. In practice, human decision-making often deviates from normative economic rationality standards: Subjective utility and expectations do not match objective gain and probabilities of a decision. Also, humans make use of intuitive heuristics and are likely to give in to emotional impulses and habits.

In his influential book “The paradox of choice”, Barry Schwartz formulates a general problem of modern (American) society: “Instead of increasing satisfaction and well-being, the abundance of choices is increasing the levels of anxiety, depression, and wasted time. With too many opportunities, the fear of missing out increases. As a results, humans tend to develop an aversion to trade-offs. The emotional stakes of decisions are raised by increasing amount of regret resulting from adverse evaluation of the decisions made.” [Schwartz and Kliban, 2004]. Schwartz also states that the evaluation of decision consequences has changed with respect to social comparison. With the appearance of social networks, this factor seems to get even more relevant.

To conclude, there is a strong demand for virtually all kinds of decision support, either in the form of a coach or counsellor, pen-and-paper techniques or modern computer systems.

### 5.1.2 Background on Decision Support

One of the first methods to support humans with rational decision-making was introduced by Benjamin Franklin [Sparks, 1844], which is often also referred to as *Pros and Cons*. Since Franklin’s early approach, the subject of decision-making has been covered by many research disciplines like psychology [Janis and Mann, 1977, Clemen, 1996, Schwartz and Kliban, 2004], game theory and business economics [Neumann and Morgenstern, 1947, Simon, 1979]. Also, a large set of guidance literature on decision-making has been published. Consulting a professional coach or counsellor is a more expensive alternative.

In the field of business economics, the discipline of decision-making methods is often associated with multiple criteria decision analysis (MCDA). In contrast to our work, the focus of MCDA lies on formal methods and mathematical models to calculate an optimal solution. While the early software solutions were based on spreadsheets, recent systems usually come with web-based interfaces. An overview of MCDA methods can be found in [Figueira et al., 2005].

## 5.2 Benjamin Franklin's Method

### 5.2.1 Pros and Cons

In a letter to his friend John Priestly dated back to 1772, Benjamin Franklin first mentioned a systematic method for rational decision-making [Sparks, 1844]:

“(...) my Way is, to divide half a Sheet of Paper by a Line into two Columns, writing over the one Pro, and over the other Con. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure. When I have thus got them all together in one View, I endeavour to estimate their respective Weights; and where I find two, one on each side, that seem equal, I strike them both out: If I find a Reason Pro equal to some two Reasons con, I strike out the three. If I judge some two Reasons Con equal to some three Reasons pro, I strike out the five; and thus proceeding I find at length where the Ballance lies; and if after a Day or two of farther Consideration nothing new that is of Importance occurs on either side, I come to a Determination accordingly.”

Franklin's method is often simply referred to as “Pros and Cons”. The basic process steps of his method can be summarized as follows:

1. Frame the problem into a clear yes/no either/or decision. question regarding a course of action.
2. Write down a list of all Pros and Cons.
3. Assess and weight Pros and Cons with respect to their importance and entrance probability.
4. Analyse and reflect on the weighted list.
5. Review and conclude.

### 5.2.2 Moral Algebra

Franklin was aware that a good rational decision can not be simply calculated by summing up weighted arguments. Instead, making a decision always involves various moral, emo-

tional or other subjective aspects that just cannot be formalized. Franklin himself described his method for decision-making as “moral algebra”:

“And tho’ the Weight of Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to take a rash Step; and in fact I have found great Advantage from this kind of Equation, in what may be called Moral or Prudential Algebra.”

Although Franklin had no knowledge of concepts such as cognitive load, limited working memory capacity, external memory or decision-making patterns, Franklin’s perspective anticipated many ideas and concepts used in modern decision theory and cognitive science. For example, he suggests to avoid impulsive responses, i.e. hypervigilant decision-making (See Sect. 5.3).

### 5.2.3 Discussion of Franklin’s Approach for Coaching

Benjamin Franklin’s idea of a moral algebra gives an important implication for decision support, and computer based decision coaching in particular: Although the decision maker might be able to create a formal representation with weights, the best solution does not necessarily derive from a maximization function. Choice is a subjective task, based on personal values and goals, which have to be evaluated carefully by the decision maker himself. Therefore, a decision support system such as VPINO does not need to suggest a particular option or decision.

From our modern perspective, Franklin’s method may seem too simplistic and outdated. Nevertheless, it is still commonly recommended by blogs, supportive literature and even professional coaches. Many decision support websites or books present templates as the example displayed in Fig. 5.1. In accordance with Franklin’s Idea of a moral algebra, the deciding human has to review the list of Pros and Cons and reflect about relevant issues for their specific problem instead of calculating a choice. However, the design of the spreadsheets, weight categories (risk, probability, importance, gain, etc.) and reflection questions can vary depending on the authors preferences. For example, the review and reflection questions include the following (quoted literally from [Nickols, 2015]):

- How do the Pros and Cons stack up against one another?
- How does the importance of the Pros compare with the importance of the Cons?
- How does the probability of the Pros compare with the probability of the Cons?
- What is the “mix” of importance and probability? What patterns do you see?
- What patterns do you see? What do they tell you? What is the overall balance of consequences?
- What are you getting? What are you giving up? Are the trade-offs worth it?

### 5.3. Human Decision-Making Patterns

***Making Decisions like Ben Franklin: A Job Aid for Decision-Making by Decision Makers***

**The Decision:** \_\_\_\_\_

PROS	I	P	W	CONS	I	P	W
Rate Importance (I) & Probability (P) on a scale from 1 to 10				Determine Weight (W) by multiplying I rating x P rating			

**Figure 5.1:** An example of common spreadsheets for decision-making with Benjamin Franklin’s method, usually provided with a set of possible review and reflection questions.

- What are you risking and are you willing to risk it?
- When do you have to decide?
- Are there any overriding factors?
- Are there any inconsistencies to be resolved?

However, in Franklin's approach, the decision maker commits to a single course of action, either for or against a particular alternative. Thereby, a decision is limited to two alternatives (for/against or yes/no). Also, Franklin's approach is driven by evaluation of a fixed set of collected alternatives only. In contrast, more recent approaches focus on goals and values related to the decision problem.

### 5.3 Human Decision-Making Patterns

Each human being has a different strategy when approaching a decision. Whereas some people decide hastily, others tend to procrastinate. Yet others assess all alternatives before making a choice.

### 5.3.1 Vigilance and Hypervigilance

Research by Janis and Mann focused on the way individuals usually make their decisions. They presented a model of decision-making in which they distinguished between *vigilant* and *hypervigilant* decision-making patterns [Janis and Mann, 1977].

Vigilant decision-making is considered the ideal pattern of decision-making: the decision maker “searches painstakingly for relevant information, assimilates information in an unbiased manner, and appraises alternatives carefully before making a choice” [Janis, 1982, p. 73].

Hypervigilant decision-making, on the other hand, represents an impulsive, disorganized pattern of decision-making: The hypervigilant decision maker’s “thought processes are disrupted ... his thinking becomes more simplistic. He is likely to search frantically for a solution, persevere in his thinking about a limited number of alternatives, and then latch onto a hastily contrived solution” ([Janis and Mann, 1977, p. 51]). An overview between a vigilant and a hypervigilant approach on decisions according to [Janis and Mann, 1977] is provided in Tab. 5.1.

Vigilant	Hypervigilant
- systematic	- non-systematic
- organized information search	- selective info search
- thorough consideration of <i>all</i> available alternatives	- consideration of <i>limited</i> alternatives
- devotion of sufficient time to evaluate options	- rapid evaluation of data
- re-examination and review of data	- no extensive review or reassessment

**Table 5.1:** Characteristics of vigilant vs. hypervigilant decision-making patterns

Whereas hypervigilance is generally viewed as a “defective coping pattern”, a vigilant, analytic pattern of decision-making is often associated with high-quality decisions. However, there has also been criticism on this proposition.

### 5.3.2 Heuristic and Intuitive Decision-Making

[Gigerenzer and Goldstein, 1996] argues that human intuition and heuristics often lead to better results than a systematic evaluation on all available details. Therefore, intuition should be also considered as an essential part of rational decision-making.

In accordance with Gigerenzer, Barry Schwartz also criticizes strictly maximization of objective values as the only way to good decisions. [Schwartz and Kliban, 2004] distinguishes between two types of decision makers, *maximizers* and *satisfiers*. Maximizers are perfectionist, highly vigilant decision makers that need to assure they took the best possible options, whereas satisfiers take the first options that satisfies their requirements. A



satisficing decision-making pattern is related to more happiness and satisfaction with a decision.

However, systematic evaluation and human intuition do not necessarily exclude each other. Obviously, Franklin's original approach supports a vigilant pattern for decision-making. Nevertheless, his "moral algebra" suggests extensive evaluation and reflection before making a decision, which does not necessarily exclude heuristic or intuitive assessment of the alternatives.

VPINO aims to assist decision makers with a structured conversation for high quality decisions. Therefore, VPINO employs a combination of Franklin's approach and the following, more goal oriented strategy as presented by [Schwartz and Kliban, 2004]:

1. Figure out your goals.
2. Evaluate importance of each goal.
3. Array the options.
4. Evaluate each of the options to meet your goals.
5. Pick the winning option.
6. Use the consequences of your choice to modify your goals, the importance you assign them, and the way you evaluate future possibilities.

For the effectiveness of a coaching session with VPINO, we expect the decision type of a user to have an influence on the outcome of the conversations. More precisely, our hypothesis is that VPINO can effectively help users with a generally more hypervigilant decision making strategy, more than users with a generally more vigilant approach on decisions.

## 5.4 Rational Decision Dialogue Structure

This section presents the adoption of Benjamin Franklin's decision method to a dialogue plan for VPINO. First, the internal structure of the decision problem and the attributes in VPINO's structured knowledge memory (SKM) are explained. This internal structure for the decision problem is filled with data step-by-step with the sub-dialogues as described in Sect. 2.5. After that, the organization of coaching questions as a dialogue plan in three stages is described.

### 5.4.1 Structured Knowledge Memory for Decision Coaching

As presented earlier in Ch. 2, VPINO keeps a structured knowledge memory, the technical correspondence to the cognitive concept of "working memory" known from cognitive science. For the purpose of rational decision coaching, the SKM is organised in a tree-like structure for the decision problem and all of its sub-aspects discussed throughout the conversation.

- The root represents the *Decision problem* as a whole.
- The descendants of the root are the individual *Options*.
- Each option has two types of descendants: the *Pro arguments* and the *Con arguments*.

Each of the nodes presented in the following list holds a set of attributes. Throughout the conversation, the problem-tree grows dynamically and the options of the individual nodes are filled in according to VPINO's questions and the user's responses. Keep in mind that in contrast to Franklin's approach, more than two options are allowed. In order to keep the conversation focused, the number of options is limited to a maximum of five. An overview on the SKM in the context of conversation organisation can be found in Fig. 5.2.

#### 5.4.2 Attributes

The attributes are set step-by-step throughout the conversation. For each attribute, VPINO asks a specific question requesting that particular piece of information (i.e. enters the corresponding sub-dialogue). Attributes can be stored as plain text as typed in the user message, others are boolean or numeric values. After finishing a sub-dialogue, the user's answers are parsed and the information is extracted depending on the specific question, as described in Sect. 2.5.

Examples for numeric attributes are the number of possible options, or a numeric assessment of the risk as described by the user. In general, option and argument descriptions by the user are stored as plain text. They are used for reference resolution or presenting specific information later in the conversation (e.g. to show a list of all options, or a list of all arguments).

- **Attributes of the decision problem** The SKM stores the list of options and the user's general description of the problem, the client's short-term and long-term goals, and whether a compromise is possible and desired. Furthermore, we store the user's criteria for a hypothetical/theoretic optimal solution, and whether or why these criteria describe potentially realistic goals.
- **Attributes of an option** For each option, besides the list of Pro and Con arguments, the SKM stores the user's answers on the following aspects: a description of that option, a short-cut description in a few words, risk, chance and the user's personal influence on the risk.
- **Attributes of an argument** Arguments include the user's description of that argument and a boolean flag indicating whether or not this argument is rated as particularly important by the user.

Besides the information about the decision problem given by the user, the SKM is also used to store internal information about the state of the dialogue as well as a complete history of the conversation so far.

### 5.4.3 Dialogue Plan

As mentioned before, we follow the basic ideas of Benjamin Franklin. For decision-making between multiple alternatives, we have adopted the process of good decision-making as suggested by [Schwartz and Kliban, 2004] (See Sect. 5.4) into a dialogue plan for VPINO. The dialogue plan consists of the three coaching stages:

- **Stage I:** Problem framing and goal setting
  - Discuss scope and expectations of the conversation.
  - Identify goals.
- **Stage II:** Options and arguments collection
  - Collect items.
  - Weight/rate items.
  - Identify favourites.
- **Stage III:** Review and next steps
  - Pairwise option discussion.
  - Discuss potential compromise.
  - Plan next steps.

A detailed description is given in the following paragraphs. The stages are subdivided into sub-dialogues, as described in Sect. 2.5). To summarize, each sub-dialogue represents the part of the dialogue that is devoted to a specific attribute in the SKM. For example, a sub-dialogue may request the description of a new option, a Pro or Con argument, or the name of the favoured option by the user. Within each stage, particular sub-dialogues are dynamically selected by the dialogue manager, which monitors the internal state of the conversation, the attributes stored in the SKM and particular precedence constraints between sub-dialogues.

### 5.4.4 Stage I - Problem Framing and Goal Setting

The first stage begins with a greeting and a short introduction of VPINO, followed by a clarification of the conversation scope. A set of sub-dialogues corresponds to “problem framing” in Franklin’s approach. Beyond that, VPINO aim to manage the user’s expectations. More specifically, we clarify what the conversation will do and what it will not do,

especially that VPINO will not suggest any option. The user is asked for personal goals for the conversation as well as long-term and short-term goals related to the decision problem. The stage's questions cover all necessary attributes for the *Decision Problem* in the SKM.

Overall, the first stage includes 23 sub-dialogues. However, not all of them are necessarily discussed with the client, depending on the results of preceding sub-dialogues, the progress of the conversation, and the SKM (same for the sub-dialogues in the following stages).

#### 5.4.5 Stage II - Options and Arguments Collection

In the second stage, VPINO's questions ask the user to describe the set of option and the Pros and Cons for each option, step-by-step. Overall, the second stage consists of 43 sub-dialogues. The sub-dialogues include questions for all attributes of *Options* and *Arguments* in the SKM. Furthermore, the user may identify particularly important Pros and Cons. Particular sets of sub-dialogues in this stage are repeatedly used in a single session, but within the context of a different option. More precisely, all questions devoted to collecting arguments and attributes of options are repeated.

The approach suggested by Franklin rates each Pro and Con separately with regard to risks, their probability of occurrence, and importance of that argument. In contrast, the user of VPINO has to give a personal assessment of each option as a whole, instead of rating each argument separately. We did this to keep the dialogue shorter and avoid boring the user with repeated requests on every single argument.

VPINO will not give an explicit definition of risk, the interpretation of risk is left to the user. This is the right way for a system such as VPINO, which only supports the user in reflecting on her thoughts (in contrast, a system that indeed computes and suggests a decision, must come to a consensus with the user on the exact meaning of risk). Nevertheless, leaving the definition of risk to the user was criticized by some users in the study. At the end of Stage II, the user is asked to name an intuitively favoured option so far.

#### 5.4.6 Stage III - Review and Next Steps

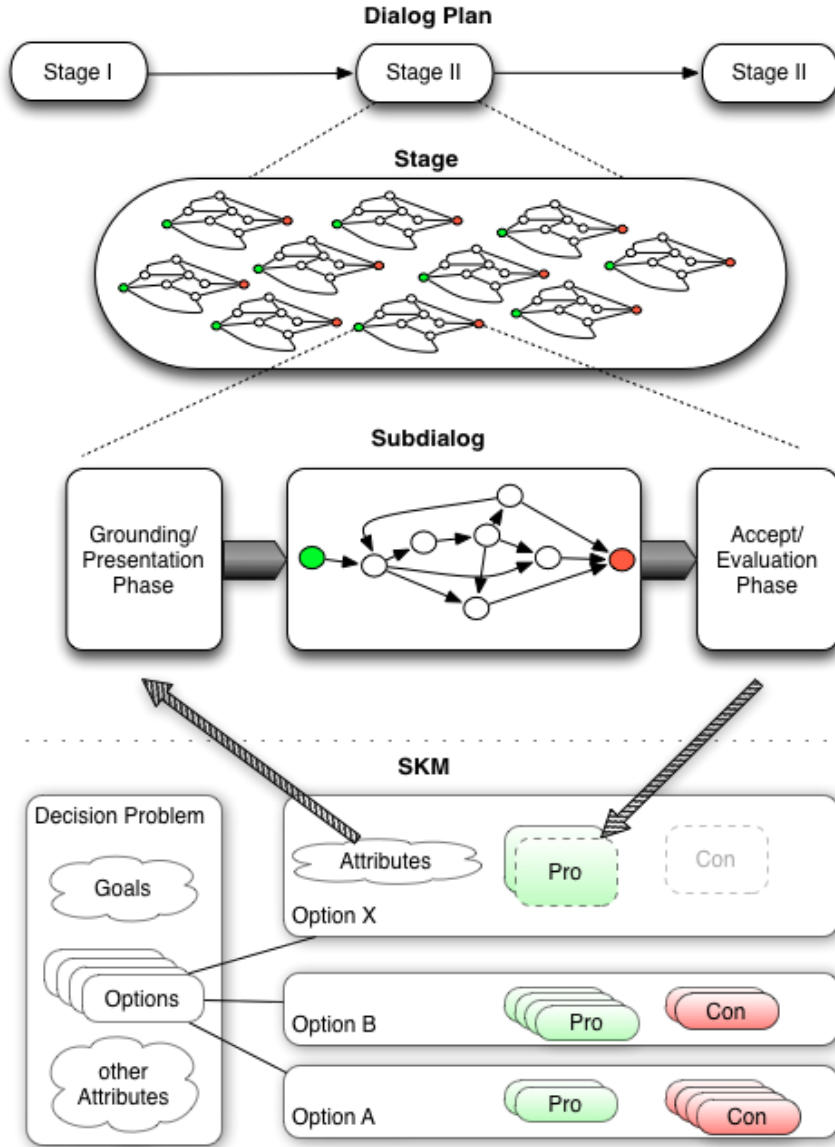
After collecting options and arguments and discussing each of them separately, the aim of Stage III is to elaborate a more differentiated view on the options by putting them into relation to each other. As mentioned before, VPINO does not suggest a solution/particular option at any point. Instead, it always tries to encourage the user to reflect about the options to come to a conclusion by herself. VPINO verbalises/paraphrases the options based on the available information in the SKM, and selects some of the options for detailed pairwise comparison (a description of pairwise option comparison follows in Sect. 5.5). Thereby, VPINO tries to reveal problematic or discussable facts the user has potentially not taken into account yet. To do so, VPINO uses a set of reflection sub-dialogues, inspired by

#### 5.4. Rational Decision Dialogue Structure

the set of reflection question as presented in Sect. 5.2.3. At the end of Stage III, the user should be able to point out a favourite option.

The user is asked to identify a favourite option and justify/reason why this option is ultimately preferred over the others. The user is asked to sum up the results and make a plan for the next steps. For some users, the next step is to collect more information as a base for a good final decisions. Others may have already made a choice and plan the steps how to put their decision into practice. Finally, VPINO finishes the conversation.

Overall, Stage III includes 25 potential sub-dialogues. Depending on the course of the conversation, an average conversation will only use about 8 of them. The remaining sub-dialogues are not suitable for particular conversations, for example planning steps on how to put an alternative into practice, when no choice has been made yet.



**Figure 5.2:** An overview of conversation organization and the SKM for rational decision coaching with VPINO. The three stages of the conversation consist of a set of sub-dialogues. A sub-dialogue is devoted to create or change a particular attribute in the SKM. The SKM is a tree-like data structure for the problem, options, arguments and all corresponding attributes. The particular sub-dialogue shown here is devoted to collecting a new pro argument. The particular sub-dialogue presented in this figure creates a new Pro argument in the SKM and stores the description text as given by the user.

## 5.5 Pairwise Option Comparison

The most important, yet most difficult part of the decision making process is review and reflection. In order to further improve the effectiveness of the decision coach, a good support to help a client with reflection and evaluation of all alternatives is required.

VPINO provides this support for the user with pairwise comparison of all options. After collaboratively collecting all the information about the options, their Pro and Con arguments and weights in the earlier stages of the conversation, VPINO determines “interesting” pairs of options that are potentially relevant for further discussion.

While pairwise comparison could be repeated for all potential combinations of options, we decided to limit the comparison to two options in our study. We did so in order to avoid bored users due to too long conversation. Most conversations in the user study did not exceed the amount of three options.

For pairwise option comparison, VPINO needs to (1) determine the *option characteristics* potentially interesting for pairwise comparison, (2) select two pairs from the set of interesting options, (3) generate *paraphrasing* for each interesting pair, and (4) select a suitable question for the selected pair of options.

In contrast to the generally shallow language understanding of following the dialogue act sequence model in the sub-dialogues, pairwise option comparison is a more complex task that involves evaluation of the information collected so far. Thus, it takes place in the evaluation phase of a sub-dialogue. The result of this evaluation task, i.e. the paraphrasing on the pair of “interesting” options, is presented in the grounding phase of a following sub-dialogue, which deals with the situation described in the paraphrasing. The following example statement by VPINO, as taken from the second user study, illustrates an actual result of option comparison process.

*In the beginning of our conversation, you said you could not point out a favourite option yet.*

*The option **Stay in Berlin** yields a high chance with a low risk.*

*On the other hand, your Cons seem to slightly overweight the pros here.*

*In contrast, option **Move to Frankfurt** seems rather risky with a low possible gain.*

*Furthermore, cons seem to overweight.*

*When comparing these two, what are the reasons to consider **Move to Frankfurt** as a reasonable solution?*

### 5.5.1 Option Characteristics

We determine “interesting” options based on a combination of the risk/chance ratio, counting arguments and whether there were more important pros or cons.

**Risk/chance ratio**

For every possible option, the client is asked for her assessment on the risk and chance of that option. Both risk and chance are stored as numeric values between 1 to 10 (where 1 is low and 10 is high).

Fixed borders for the high and low values on risk and chance would be the first thought to categorise the options. Unfortunately, some humans tend to pick values around the median, others are more likely to name extreme values such as 1 or 10. Due to this experience, a variable border  $db \in \{2, 3\}$  is used, depending on the dispersion of the user's answers. The target categories for the risk( $r$ )/chance( $c$ ) ratio are defined as follows:

1. High Risk / High Chance (*hi/hi*) when risk and chance are less than 7
2. Balanced (*bal*) when risk and chance are less than 4 or the difference is lower than the user specific border value  $|risk - chance| > db$
3. High Risk / Low Chance (*hi/lo*) when  $r - c > db$ , unless one of the categories *hi/hi* or *bal* apply.
4. Low Risk / High Chance (*lo/hi*) when  $c - r > db$ , unless one of the categories *hi/hi* or *bal* apply.
5. Low Risk / Low Chance *lo/lo*

**Weights of arguments**

Apart from analysing the risk/chance ratio, weights of the pro and con arguments are categorised. VPINO counts the pro and con arguments and applies one of the following categories:

1. *Strong bias pro! (resp. con!)*: user has named more pros than cons *and* additionally has identified one or more pros as important.
2. *Slight bias pro (resp. con)*: user has named more pros than cons *and* additionally has not identified one or more pros as important.
3. *Neutral*: either both pros and cons do have important arguments *or* both do not and additionally have an equal number of arguments.

VPINO calculates a positive, negative, neutral or uninteresting rating of that option as shown in Tab. 5.2. As a result, options with a very high chance rating by the client are always tagged as interesting for further discussion, as well as options with a strong overweight on pros and a particular import pro (same for cons). Options rated as negative or positive refer to "obvious" option characteristics. For example, an option with a high risk and a strong overweight on cons is obviously not the best choice. Nevertheless, it was listed by the user and therefore makes a candidate for further discussion.



	risk / chance			
	hi/hi	hi/lo	lo/hi	balanced
pro!	+	o	+	+
pro	o		+	
con	o	-	o	
con!	-	-	o	-
neutral	o		o	

**Table 5.2:** Combinations for interesting options based on risk/chance ratio and overweight on arguments ("!" denotes strong overweight). Resulting in positive, negative and neutral positions. No entry for uninteresting option.

### 5.5.2 Option pair selection

From the set of options evaluated as positive, negative or neutral-interesting, a number of two pairs maximum are chosen for further discussion in the conversation. The option denoted as the client's favourite is added to the list in case it was not yet chosen as a candidate for the list of interesting options  $O$ . From the list of interesting options  $O$ , VPINO builds a list of all possible pairs of interesting options  $(o_1, o_2) \in P$  for  $o_1, o_2 \in O$  and  $o_1 \neq o_2$ .

From this list, VPINO selects two pairs of options: Highest priority have pairs, including the client's favourite option. Pairs that include options with positive or negative ratings are preferred over pairs including neutral-interesting options.

### 5.5.3 Paraphrasing Option Pairs

We use a set of preformulated text snippets to generate the paraphrasing for the selected option pair. Each text snippet is devoted to describe a single fact of an option. Depending on the selected pair of options, VPINO uses linking words that either contrast or compare the facts of the options. VPINO does so by connecting the statement's facts with openings such as "*On the other hand*", "*Furthermore*", "*Also*", or "*In contrast*", depending on the characterization as a positive or negative fact. For example, VPINO would contrast a long list of pros for the first option to a long list of cons for the second option.

However, although pointing out or even contrasting obviously positive or negative options, VPINO will never suggest one over the other. Instead, discussion and explanation why the client would possibly prefer one option over another, is left to the client.

Given the case VPINO did not identify a single interesting option at all, it selects a random pair of options (still preferably including the favourite). VPINO points out random facts about the options, as in the following example:

*In the beginning, you identified option “**Keep the old one...**” as your favourite.  
The cons seem to slightly overweight the pros here.  
Option “**second hand**” looks a bit risky.*

Depending on the pair of interesting options selected for further discussion, a sub-dialogue is chosen. For example the client is asked to justify his previous decision (in case of a favourite option) or to formulate his general thoughts on comparing these two options, for example “*When comparing these two, what are the reasons you prefer ‘**Keep the old one**’?*” or “*What are the reasons to still consider the first/the second option as good solution?*”

## 5.6 Semantic Reference Resolution on Options

Throughout the conversation, VPINO repeatedly faces the common situation where the user has to refer to one of the collected options or arguments, as in the example presented earlier in this work in Fig. 2.4. More specifically, when asked to name a favourite or particular important option, the user somehow references a particular option by giving a detailed description, a brief shortcut, the index or position of an option. This reference has to be resolved by VPINO through parsing the user answer.

The user study on decision coaching revealed the need for elaborate strategies on such conversation related problems, which are trivial for humans, but hard for computers. Failure in such a trivial tasks leads to disappointment and thereby ineffective conversation. The method of semantic reference resolution presented in this section is the final version as used in the second user study in Ch. 7. It the result of continuous improvement and the insights gained in the first user study in Ch. 6

The task of semantic object reference resolution is not to be confused with coreference resolution, a well studied problem in linguistics. A coreference is defined as the occurrence of two or more expressions in a text that refer to the same entity or person. For example, “Daniel thought that he could achieve anything” is generating a coreference between “Daniel” and “he”, which refer to the same person. In contrast, semantic reference resolution of objects, i.e. particular options and arguments in a fixed setting in chat conversation is a different task with a completely different focus. The problem of Option and Argument resolution in VPINO can be described as follows:

Given a fixed, ordered set of objects represented by a textual descriptions, and an input text string, the output of a semantic reference resolver is the index of the object semantically described by the input string. In the following, we describe the strategy developed through the course of this work.

### 5.6.1 Baseline approach

The implementation of VPINO as used in the first user study, applied a simple straight forward approach: VPINO calculates a numeric text similarity value for each option description and the query text. It chooses the option with the highest similarity value between the query string and the given descriptions that passed a particular threshold.

However, this approach turned out to be quite ineffective, especially when the length of input and description texts differed. For example, some Pro or Con arguments were described by a single word, whereas the user referenced these arguments with a whole sentence. In order to improve this baseline approach, we evaluated several different approaches for semantic similarity resolution.

### 5.6.2 Normalization

We evaluated normalization by spell-correction, stemming and lemmatization before calculating text-similarity measures. Unfortunately, normalization largely decreased both precision and recall of the similarity measures on our test-set. Especially spell-correction turned out to be a disimprovement. That is because both components were trained on written text corpora, different from the sloppy chat language as commonly used in conversation with VPINO. Consequently, spell-correction often resulted in a replacement of sloppy chat language with more or less similar “real” words.

### 5.6.3 Semantic Similarity Measures

An approach commonly used for text categorisation and semantic reference resolution is counting the N-gram overlap [Lin, 2004]. A word N-gram is a sub-sequence of N words, from a given sequence of text. Reoccurring pairs or triplets of words (Bi/Tri-grams) are a hint for similar content of documents. We evaluated an approach based on stemming the words and counting word Uni-gram and Bi-gram overlaps. However, counting N-Grams turned out to be ineffective for the very short text descriptions given by the users. Tri-grams matches rarely occurred, Bi-grams overlap was also sparse and did not improve quality of the results on most occurrences.

Some approaches on measuring semantic distance between two texts are statistical methods based on distributional similarity and Latent Semantic Analysis [Han et al., 2013]. Latent Semantics Analysis is a technique for comparing texts using a vector-based representation that is learned from a corpus [Landauer and Dumais, 1997]. LSA works fine on larger documents. However, the problem for specific task of identifying a particular description from a set of options is, that most of these option descriptions themselves are highly semantically related, with only minor differences. Therefore, each of the options potentially produces a high semantic similarity rating. As a result, common similarity measures resulted in a large amount of false positives on our qualitative test set.

Since many of the similarity measures were not suitable for reference resolution in chat conversation, we used a combination of several approaches using a majority vote strategy. Majority vote is a decision strategy, where one solution out of a set of choices is determined as the one with more occurrences than any other.

- In a first step, prior to the application of the majority vote, the user utterances are parsed for positional references, such as “*the first*”, “*the last*”, “*the latter*” ...), numeric identifiers for the index, and quantitative identifiers (“*none*”, “*all*” ... ) by matching key-phrases from a list of keywords.
- In a second step, VPINO applies a majority vote based on the following components. VPINO calculates the text similarity between the option/argument description and the user’s answer text, as used with the baseline approach. The option with the highest similarity that passes a certain threshold receives the vote.

In German language, a statement can often be reformulated with an inverse order (at least partially) of words. Therefore, the sequence of word tokens from the query text is inverted, and again evaluated with the text-similarity measure. The option with the highest similarity that passes a certain threshold receives the vote. This additional word sequence inversion strategy led to a large improvement.

Finally, the option with the highest word overlap on the query text receives a vote.

#### 5.6.4 Evaluation

The reference resolution was developed and optimized for a qualitative evaluation test set. Since the test-set is manually constructed we do not provide a quantitative evaluation here. Moreover, it is used to find a solution that is most suitable in practice and detect particular hard cases. The evaluation set is based on real examples collected in the first user study on decision coaching. A test case is a pair of an option set and a query text. Test cases use 30 option-sets, selected by size, quality of content, and representativeness. Test cases consist of 40 user references collected in the user study, added up with a number of 200 “artificial” references authored by the developers. The test set was designed to cover a large variety of references as expected in practice.

A major objective was to avoid false positives, since it is hard to correct these mistakes in the conversation. However, in case no reference could be resolved, VPINO asked for a particular number or the exact shortcut description, as stored in the SKM. The method for object reference resolution presented in this section worked very well on our test set, but also in practice, especially in the second user study.

## CHAPTER 6

# Rational Decision Coaching: User Study

---

In this chapter, we present a user study on the effectiveness and usefulness of VPINO as a decision coach. We evaluated effectiveness and personal influences on success of a coaching session with VPINO. The results of the study showed that serious, motivated users evaluated the acceptance and usefulness of VPINO as a decision coach quite positively. Participants that are more open to computer based decision support held better and more fruitful conversations than those with a sceptical attitude. This quantitative result conforms well to our qualitative observation that VPINO shows good human like behaviour whenever the user is cooperative and motivated. We also found that users with a more hypervigilant approach to decisions particularly benefit from VPINO. The results of this chapter have been published in [Mäurer and Weihe, 2015].

The the design of the user study and the measures used are described in Sect. 6.1, followed by the results of the study in Sect. 6.2 and Sect. 6.3. Finally, we sum up the results and conclusions in Sect. 6.4

## 6.1 Setting

### 6.1.1 Study Design

We conducted a study to evaluate user acceptance and effectiveness of the VPINO decision coach. Our study was conducted as a field experiment with 129 (29.5% female) participants who voluntarily signed up. On average, participants were 25.14 years old ( $SD = 6.04$ ). The study was conducted via our web-interface, which allowed the participants to choose time and location freely. Study took part in single sessions, which were split into three subsequent phases:

- In a multiple choice *pre-conversation survey* at the beginning of the session, we tested for the participant's approach on decisions (vigilant/hypervigilant), the Big Five personality traits and the participants' attitude towards decision coaching and dialogue systems (chatbots) in general.
- In the following conversation phase, participants were instructed to hold a chat-conversation with VPINO, on a rational decision problem of their own choice.
- The conversation phase was immediately followed by a *post-conversation survey* to evaluate effectiveness of the conversation on the participant's decision problem, general user acceptance and self-perception of the participant's own work and willingness to cooperate with the dialogue system.

Participants were provided with a 10€ incentive for successful participation in all three phases.

### 6.1.2 Measures

**Pre conversation** The participant's decision type was tested with the Melbourne Decision Making Questionnaire [Mann et al., 1997] on a five-point Likert scale (0 = disagree strongly to 4 = agree strongly): Whenever I face a difficult decision ...

- I feel pessimistic about finding a good solution.
- I feel as if I am under tremendous time pressure when making decisions.
- The possibility that some small thing might go wrong causes me to swing abruptly in my preference.
- I cannot think straight if I have to make a decision in a hurry.
- After a decision is made I spend a lot of time convincing myself it was correct.

We tested the Big 5 personality traits openness, conscientiousness, extraversion, agreeableness and neuroticism with the Big 5 inventory by [Rammstedt and John, 2007].

The attitude (*atti*) towards computer based decision coaching was measured by a scale based on 5 items (Cronbach's  $\alpha = .744$ ) on a five-point Likert scale (0 = disagree strongly to 4 = agree strongly):

- A chatbot could help me with solving my problems in general.
- A chatbot could help me with *personal* problems.
- A chatbot could help me with *decision* problems.
- I would discuss private topics with a chatbot.
- I would discuss career-related topics with a chatbot.

The term “chatbot” was used in the survey-questions because it is the most common description for chat-based artificial dialogue systems. Nevertheless, a detailed explanation for term “chatbot” was provided in the instructions of the survey.

**Post conversation** For the evaluation of the VPINO as a professional coach, measures for effectiveness on decision-making (*decision*) were tested in a *post-conversation survey*, right after participants finished the conversation. Effectiveness on decision-making was measured by 7 items (Cronbach's  $\alpha = .913$ ) on a five-point Likert scale (0 = disagree strongly to 4 = agree strongly):

- I have reached my goal.
- I have made a decision.
- I realised what next steps I have to take.
- I have have gained clarity on my situation.
- I feel emotional relieved.
- I feel satisfied.
- The dialogue was motivating me to work on my problem.

The participants' self-perceived work with VPINO and participants' cooperativeness (*spwc*) was measured by 3 items (Cronbach's  $\alpha = .714$ ) on a five-point Likert scale (0 = disagree strongly to 4 = agree strongly):

- I responded the same way I would have with a human chat partner.

- I was respectful.
- I am satisfied with my own work during the conversation.

The general overall evaluation of user acceptance and usefulness (*overall*) were measured by 5 items (Cronbach's  $\alpha = .923$ ) on a five-point Likert scale (0 = disagree strongly to 4 = agree strongly):

- Working with the decision helper was fun.
- The conversation was motivating.
- I was positively surprised by the dialogue system.
- I would recommend the conversation with VPINO to others.
- How would you rate your overall experience with the dialogue system? (0 = bad 4 = very good).

## 6.2 Results

The average length of the *chat* dialogue was about 23:15 minutes (SD = 695 sec) with an average amount of 55.65 (SD = 15) user messages sent. Overall, users evaluated our system quite positively. Tab. 6.1 shows the amount of ratings for the most relevant evaluation questions.

		0	1	2	3	4	avg
	a	8.5	20.2	31.8	24.8	14.7	2.17
	b	14.0	11.6	19.4	37.2	17.8	2.33
All	c	10.1	19.4	21.7	35.7	13.2	2.22
N=128	d	11.6	24.8	18.6	35.7	9.3	2.06
	e	13.2	18.6	26.4	27.1	14.7	2.12
	f	5.4	14.0	29.5	41.9	9.3	2.36

**Table 6.1:** Results on most relevant evaluation questions after conversation with VPINO, answer freq. in %. Questions on the decision process (0 = disagree strongly to 4 = agree strongly): Thanks to the conversation (a) ... *I have reached my goal*, (b) ... *I realised what next steps I have to take*, (c) ... *I have have gained clarity on my situation*. Questions on overall evaluation: (d) *The conversation was motivating*, (e) *I would recommend the decision dialogue to others*, (f) *Overall user experience* (0 = bad to 4 = very good)

Average result on the question on the overall experience with the dialogue system (a) was 2.36 (SD=1.01). 51.2% of the participants evaluated their experience with VPINO as



"very good" or "good", compared to 29.5% "acceptable" and 19.4% "not so good" or "bad". VPINO also evaluated quite positively on the other relevant items: 39.5% users agreed to have reached their goal (b) and 55% realised what their next steps should be (c). 48% agreed to have gained clarity on their situation (d). The evaluation on how motivating the system was (e) is rather neutral. 41.8% of the participants would recommend VPINO to others compared to 31.8% that would rather not.

As expected, talking to VPINO did not work equally well for all participants. With our study, we also want to find out for which users VPINO worked particularly well. We investigated intercorrelations for the measures from the *pre-conversation survey* and the *post-conversation survey*. The most relevant results are shown in Tab. 6.2.

	attitude	hypervigilance	spwc
decision	.441	.274	.348
overall	.335	.322	.553
combined	.419	.319	.480

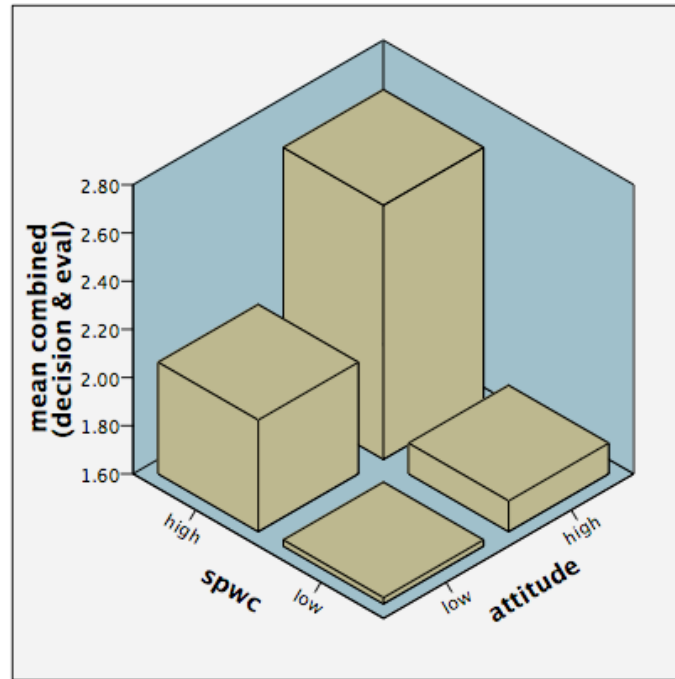
**Table 6.2:** Pearson correlations for evaluation measures ( $p < .001$  for all entries).

### 6.2.1 Openness and Attitude

In general, the user's Big Five personality traits did not seem to have influence on success of the conversation. In contrast to [Mäurer et al., 2013], no relation between openness as a general personality trait and successful conversations was found. However, we did find a relation for the user's attitude/openness on conversational agents in particular. More specifically, participants who believed that dialogue systems can support them in making a decision were more likely to reach their goals (*decision*:.441,  $p < .001$ ) and also evaluated the system more positively (*overall*:.335,  $p < .001$ ). Besides that, our results also suggest that participants with a higher rating on their self-perceived work and cooperation with the system were more likely to have higher rating for overall evaluation (.553,  $p < .001$ ) and their decision efficiency (.348,  $p < .001$ ). Fig. 6.1 displays the overall evaluation results (*overall*) in relation to self-perceived work (*spwc*) and the participant's attitude (*atti*). Participants that were more cooperative and have a more open attitude to dialogue systems evaluated significantly higher than participants that lack one of them or both.

### 6.2.2 Hypervigilance

Apart from the user's attitude and motivation, we evaluated whether the user's usual approach on decisions is related to success of VPINO. We found that users with a more hypervigilant approach on decisions tend to profit more from using VPINO than others (.319,  $p < .001$ ). Fig. 7.1 shows the difference in mean results on overall rating for users



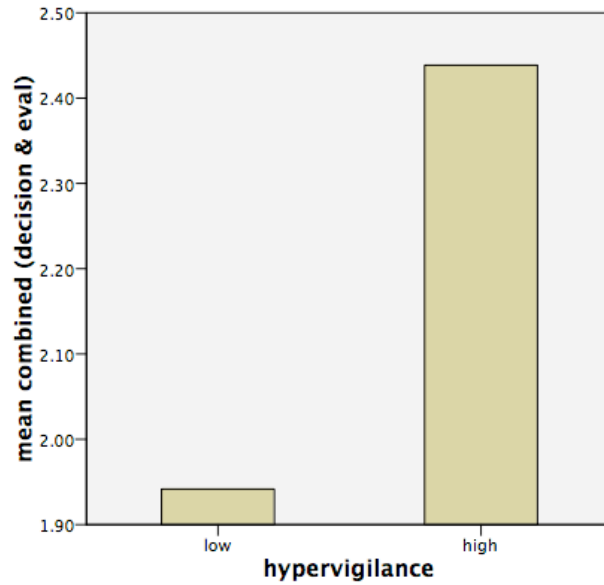
**Figure 6.1:** Influence of attitude and cooperation combined *overall* and *decision* evaluation results. Attitude towards dialogue systems (*attitude*, 1 = high, 0 = low) and self-perceived work and cooperation (*spwc*, 1 = high, 0 = low) grouped by median split.

with high and low hypervigilance (split by median). Apart from hypervigilance, both buck-passing and procrastination did not show significant effects on effectiveness. This may be due to the fact that with taking part in our study and using VPINO, participants were yet past the point of avoiding behaviour when dealing with decisions, but this is a mere speculation.

### 6.3 Qualitative Evaluation

Qualitative analysis by evaluation of the conversation transcripts confirms the statistical results presented in Sect. 6.2. In fact, our general impression was that good human-like conversations were more likely when the users were motivated, cooperative and followed VPINO's lead. The best results were obtained when users responded with precisely formulated answers on the exact question.

For a qualitative evaluation, we define a successful conversation as a natural-looking dialogue with human-like behaviour at all times. Conversations with only minor problems that did not affect den further progress of the conversation are also considered as successful. A conversation failed, if a problem either caused the participant to end the con-



**Figure 6.2:** Mean *overall* evaluation results for participants with high(1) and low(0) hypervigilance (split by median).

versation, or somehow affected the conversation in a way that effective decision support was no longer possible.

**User behaviour** While participants with successful conversations mostly followed VPINO’s lead, a few failures resulted from challenging or uncooperative users. Due to our observation, for many of them, the motivation to use our system was curiosity and challenging VPINO rather than trying to solve a decision problem.

### 6.3.1 Reference Resolution

Reference resolution problems occurred, when finishing or continuing the conversation properly was impossible due an incapability to resolve references (e.g. a user’s attempt to correct a previous answer, jumping back to an earlier question or a failure in detection of the correct reference to a favourite option). Most of these problems resulted from participants’ attempts to revise answers, or more often, attempts to add information to statements in previous turns. In contrast to reference resolution failures, most of the problems resulting from counter questions could be smoothed out by VPINO throughout the conversation, so a successful conversation was finally possible despite the error.

### 6.3.2 Dialogue Act Classification

Dialogue act classification was only a minor source of problems. Critical misclassifications, which could lead to problematic misunderstanding, for example classifying an agreement of a disagreement or vice versa, did not occur. Thanks to the sub-dialogue model, un-critical misclassifications, for example dialogue acts that lead to unexpected states in the DASM, could be handled by repair dialogues. Another interesting observation was the fact that users tend to avoid a clear no when disagreeing. While agreement is uttered straight forward and directly, most of the disagreements contained modifiers that weakened their disagreement. As a result, our dialogue act classifier did not classify some of the disagreements correctly and VPINO had to pump for more information or verify the information on simple yes/no-questions: (*"So does that mean you agree?"*).

### 6.3.3 Expectations and Disappointment

A rather small group of participants stood out by writing extremely detailed messages with three up to ten sentences per message. Due to its shallow natural language understanding capability, VPINO was not able to respond properly to detailed illustration of participants' problems. As a result, these participants got disappointed and demotivated. Some of the users accepted the limitations of the system and the conversation still led to positive results, while other users, although originally motivated, stopped their efforts on continuing the dialogue in a serious and motivated way. Once these users encountered critical situations where VPINO could not meet their expectations, they radically changed their answering behaviour to minimal answers or even stopped the conversation.

On the other hand, in many of the almost human-like conversations, VPINO was able to surprise or impress the user. A recurring example for impressed users is successful option reference resolution. Another example are particular clever pro-actively formulated responses by VPINO, resulting from precisely formulated questions and correctly anticipated user reactions.

## 6.4 Discussion and Conclusions

Rational decision-making support with VPINO could effectively help users with their decision problem. We could prove our assumption that detailed understanding is not necessarily required for the task of decision coaching, yet hold a human-like conversation. With our study we could show that highly cooperated, motivated, and serious user had the highest rate of success. Besides, users with a more hypervigilant, less systematic strategy on making decisions particularly profit from using the system.

User expectations on the intelligence, behaviour and natural language understanding capabilities of VPINO seem to have an influence on user acceptance and effectiveness of the conversation. While intelligent behaviour by VPINO could impress users and motivate, a lack of intelligent behaviour leads to decreased user acceptance, and therefore less cooperative users. Whereas classical chatbots try to overcome this problem with obfuscation tricks to simulate intelligence/cleverness, (for example switching the topic or making a joke), systems for professional use do not have that option.

For the professional scenario of decision coaching, future work will require further improvements of solving intelligent sub-tasks that are relevant for the ongoing conversation and are helpful/have a value for the client. Furthermore, future work needs to evaluate the usefulness of rational decision support with VPINO on a more broad target group of users.

## CHAPTER 6. RATIONAL DECISION COACHING: USER STUDY

## CHAPTER 7

# Rational Decision Coaching: Follow-up Study

---

We conducted a follow-up study on rational decision coaching with an improved version of VPINO. The goal of this study was to gain further insights by focusing on a broader target group with respect to age, gender distribution, and educational background. Furthermore, we examined the effects on the participant's emotional assessment of their specific decision problem. For the user study, the set of sub-dialogues was optimized and particular natural language understanding tasks were improved.

**This chapter is structured as follows:** The motivation for this user-study and a brief description of the improvements on VPINO is presented in Sect. 7.1. The setting of the user-study is described in Sect. 7.2. The results of the study are presented in Sect. 7.3, while user feedback is presented in Sect. 7.4. The chapter closes with a discussion of the findings and conclusions in Sect. 7.5.

## 7.1 Motivation and Improvements on VPINO

VPINO is intended as a highly available coach to support a large number of users with their problem. In order to evaluate VPINO’s effectiveness as a decision coach on a broader target group of participants, we conducted a follow-up user-study. Therefore, the participants were recruited in public places, i.e. a shopping mall and a major train station.

VPINO was improved based on the insights gained from the previous study on decision coaching (Chapt. 5). Therefore, the user feedback and the conversation transcripts from the first study were evaluated. We identified particular coaching questions/sub-dialogues that did not provide additional value for a majority of users. User feedback revealed that most participants valued the reflection part of the conversation more than a detailed discussion of the goals and a theoretical optimal solution in the problem framing phase at the beginning of the conversations. Furthermore, some participants criticized the length of the conversation before getting to an interesting part, which led to shorter and less reflected responses in the important reflection phase. To straighten the conversation and make it a bit shorter, we removed a small subset of sub-dialogues from the problem framing phase. The removed coaching-questions were originally devoted to identify theoretically ideal solutions for the user problems, that are generally unrealistic. Additionally, sub-dialogues on “changing the perspective to an external person” were removed, since these questions were often not considered helpful for the rational decision problems.

Besides optimization of the set of coaching questions, we improved the option reference resolution mechanism, and developed a more elaborate strategy for pairwise option comparison (as presented in their final version in Sect. 5.6 and 5.5).

Furthermore, the improvements with VPINO included minor changes in formulations, i.e. more precisely formulated questions and system responses, and slight improvements on particular DASM. Besides, the dialogue act classification rules were improved based on misclassified examples from the first phase.

## 7.2 Setting

### 7.2.1 Study Design

This second study on decision coaching with VPINO was conducted as a field experiment with 165 (63 female) participants, who voluntarily signed up. Participants were offered a 10 € incentive for successful participation. Apart from variations in the conversation with VPINO and slight changes on the survey questionnaires, the design of the study was identical with the first user study. Again, the study was conducted via the VPINO web-interface, which allowed the participants to choose time and location of their participation freely. The participants were instructed to answer questions in a *pre-conversation survey*, followed by the coaching phase and a *post-conversation survey* evaluation.



### 7.2.2 Measures

**Pre conversation** In the *pre-conversation survey*, we tested for the participant's approach to decisions (hypervigilance), their attitude towards decision coaching dialogue systems (*atti*), and about the nature of the decision problem.

Decision type (i.e. hypervigilance) of the participants were tested with the Melbourne Decision Making Questionnaire [Mann et al., 1997], identically to the first user study (See Sect. 6.1.2 for measures). The attitude (*atti*) towards computer based decision coaching was also reused the items from the previous study (Cronbach's  $\alpha = .789$ ).

**Post conversation** After the *pre-conversation survey*, participants held conversation with VPINO, followed by a evaluation post-test. The evaluation survey included the following measures:

- The user's work and cooperation was measured by 3 items (Cronbach's  $\alpha = .693$ ) reused from the first study and aggregated to a single work and cooperation score (*spwc*).
- In addition to the measures known from the previous study, we tested emotional relief after using the chatcoach with three items (Cronbach's  $\alpha = .853$ ) aggregated to a single *emotion* score.
  - After working with the chatcoach I feel emotional relieved.
  - After working with the chatcoach I feel satisfied.
  - After working with the chatcoach I feel less stressed.
- The effectiveness of VPINO on decision-making was measured by 4 items (Cronbach's  $\alpha = .845$ ) aggregated to a single *decision* score.
  - I have reached my goal.
  - I have made a decision.
  - I realised what next steps I have to take.
  - I have have gained clarity on my situation.
- The general overall evaluation of user acceptance and usefulness were measured by the 5 items used in the first study (Cronbach's  $\alpha = .919$ ), aggregated to a single *overall*.
- All items in *emotion*, *decision* and *overall* were aggregated into a *combined* evaluation score (Cronbach's  $\alpha = .932$ ).

### 7.3 Results

On average, the participants ( $N = 165$ , 63 female) were 29 years old ( $SD = 10.02$ ). Compared to an average age of 25 Years ( $SD = 6.04$ ) and in our first study (See Sect. 6.2), we were able to recruit a broader range of participants with respect to age, gender, and educational background.

The average conversation length with VPINO was 18:07 minutes ( $SD = 463$  sec) with an average amount of 50.31 ( $SD = 14$ ) user messages sent. Since some of the coaching-questions from the first part of the dialogue plan were removed, conversations held with VPINO were shorter than in the first user study.

		0	1	2	3	4	avg
User Study 2 N=165	a	11.5	23.6	20.0	33.9	10.9	2.09
	b	13.3	17.6	18.2	40.0	10.9	2.18
	c	10.9	13.9	19.4	43.0	12.7	2.33
	d	8.5	18.8	29.1	35.8	7.9	2.16
	e	17.0	17.0	23.6	27.9	14.5	2.06
	f	6.7	13.3	32.1	42.4	5.5	2.27
User Study 1 N=128	a	8.5	20.2	31.8	24.8	14.7	2.17
	b	14.0	11.6	19.4	37.2	17.8	2.33
	c	10.1	19.4	21.7	35.7	13.2	2.22
	d	11.6	24.8	18.6	35.7	9.3	2.06
	e	13.2	18.6	26.4	27.1	14.7	2.12
	f	5.4	14.0	29.5	41.9	9.3	2.36

**Table 7.1:** Results on most relevant evaluation questions after conversation with VPINO in this user study (Study 2), compared to evaluation in the previous user study (Study 1). See Ch. 6. Answer frequencies in %. Questions on the decision process (0 = disagree strongly to 4 = agree strongly): Thanks to the conversation ... (a) ... *I have reached my goal*, (b) ... *I realised what next steps I have to take*, (c) ... *I have have gained clarity on my situation*. Questions on overall evaluation: (d) *The conversation was motivating*, (e) *I would recommend the decision dialogue to others*, (f) *Overall user experience* (0 = bad to 4 = very good)

The evaluation results of this second user study confirmed the results from our previous study. The answer frequencies on the most interesting evaluation questions are provided in Tab. 7.1. Whereas the overall rating of VPINO was slightly less positive than in the first round (f), VPINO could slightly improve its performance with clarifying their situation (c). However, these positive effects may be due to the improvements in the reflection phase, including the advanced pairwise option comparison. Additionally, the conversation with VPINO was considered more motivating (d).

Furthermore, the qualitative results on attitude, cooperation and hypervigilance validate the results from the first users study could be validated. Pearson correlations between the measures for *spwc*, *atti*, *hypervigilance* and the evaluation results are presented in Tab. 7.2. The following sections will discuss the effects on the evaluation results.

	attitude	hypervigilance	spwc
emotion	.201*	.245*	.421*
decision	.301*	.156 <sup>T</sup>	.446*
overall	.345*	.184 <sup>T</sup>	.580*
combined	.353*	.204*	.562*

**Table 7.2:** Pearson correlations for evaluation measures as described in Sect. 7.2.2.

<sup>T</sup> :  $p < .05$ , \*:  $p < .01$

### 7.3.1 Motivation, Cooperation and Attitude

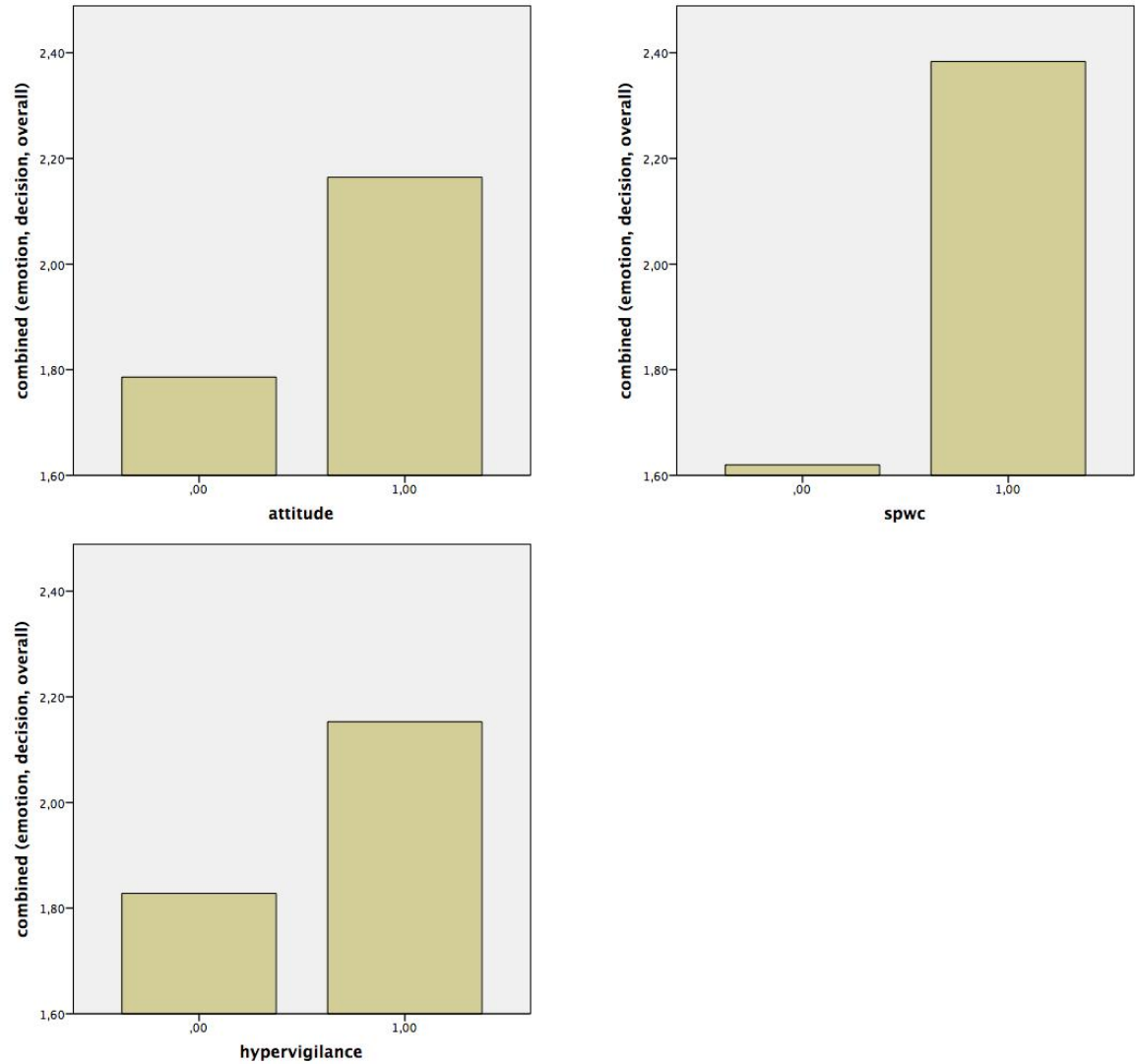
A high degree of motivation seems the best prerequisite for an effective coaching conversation. Participants with a higher willingness to cooperate with VPINO were more likely to get help with their actual problem (.446,  $p < .01$ ) and evaluated the coaching overall better (.580,  $p < .01$ ). However, the effect of cooperation is even a bit stronger than in the first user study. Still, more open and cooperative participants evaluated successful conversation significantly higher than participants, which lack one of them or both. Like in the first study, participants that are more open to computer based coaching were more likely to reach their goals (*decision*: .301,  $p < .01$ ) and evaluated VPINO more positively (*overall*: .345,  $p < .01$ ).

### 7.3.2 Hypervigilance

The results of the first user study suggested that users with a more hypervigilant approach on decisions tend to profit more from using VPINO than others. Fig. 7.1 shows the difference in mean results on overall rating for users with high and low hypervigilance (split by median).

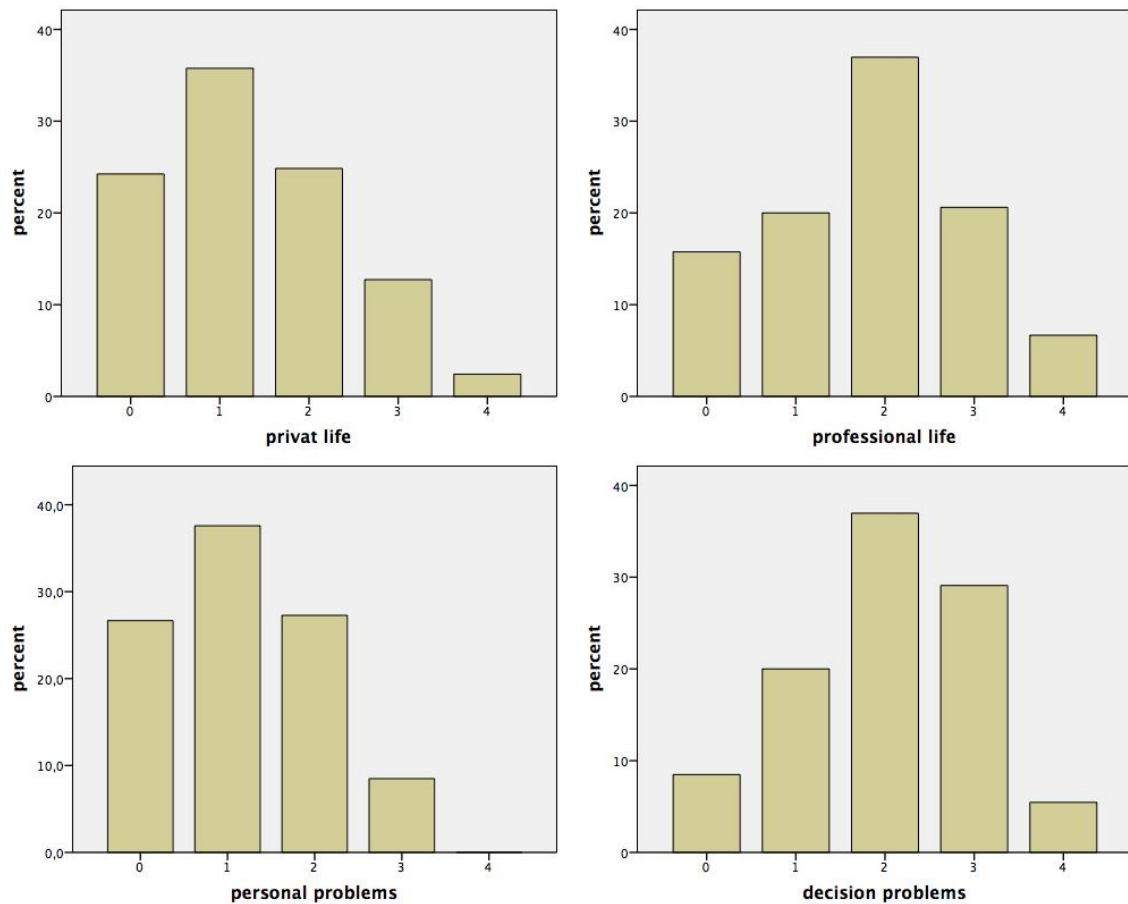
### 7.3.3 Personal vs. Professional Problems

Whereas users remain sceptical towards talking to virtual humans about personal concerns, they seem to be more open when it comes to professional concerns. Fig. 7.2 presents a chart for answer frequencies of participants' expectations and attitude towards using a computer based coach. When asked whether or not they would use a computer based chat coach, participants preferred discussing problems from their professional life (avg = 1.86, SD = 1.13) over private/personal problems (avg = 1.32, SD = 1.05). Whereas the



**Figure 7.1:** Mean evaluation results for *combined* measure. Participants grouped by the attitude towards dialogue systems (*attitude*), self-perceived work and cooperation (*spwc*), and hypervigilant approach on decisions (*hypervigilance*) (groups split by median, 1 = high, 0 = low).

participants do not expect a chat coach to help them with personal problems (avg = 1.16, SD = 0.92), they are more optimistic that a chat coach can help them with decision problems (avg = 2.03, SD = 1.03).



**Figure 7.2:** Attitude towards using a chat coach. User responses on the questions, whether willingness to discuss private/personal concerns (*private life*), professional/career related concerns (*professional life*), and whether they expect a chat coach to help them with personal problems (*personal problems*) and decision problems (*decision problems*). 0 = disagree strongly to 4 = agree strongly.

The user expectations confirm our qualitative observation that decision coaching with VPINO on problems with a higher degree of emotional pressure was generally less effective. Obviously, that effect is due to VPINO's preformulated responses: VPINO suffers from an inherent lack of empathy, no matter how friendly and polite the systems responses are authored.

For some sceptical users, conversation with VPINO was a self-fulfilling prophecy: By challenging the natural language understanding and turn-taking behaviour of VPINO, some users intentionally provoked problems. Examples for challenging user behaviour are the use of rhetorical questions or sarcastic responses.

### 7.3.4 Age and Gender

We did not observe quantitative effects for neither participants' age or gender. However, from the review of transcripts we observed a tendency that the best, most human-like conversations were produced by older participants. Younger participants tend to give short answers with a few words only. However, we speculate that this may result from differences in experience and habits with chat-communication in general.

Female participants seemed to take conversations more seriously. Almost all conversations with improper problem-subject or user behaviour of the conversation were produced by men. However, this was a general observation throughout all of our studies and is probably not related to chat coaching with VPINO.

## 7.4 User Feedback

The users were asked for positive and negative feedback, right after holding the conversation with VPINO. In the following, a list of the user feedback is presented, selected with respect to relevance and occurrence. The list of feedback is left uncommented. However, a discussion of some points criticized by the users will follow in the next section. Statements are translated from German.

### 7.4.1 Positive Feedback

- The conversation spent a little solace. The bot is refreshingly friendly.
- I could get a more objective perspective on my current situation.
- It was nice to talk with someone that was 1. unbiased 2. objective and 3. always available for consultation.
- The computer evaluated nearly all of my input correctly. Interaction was intuitive.
- Quite human like!
- I enjoyed the systematic, structured approach. Very rational.
- The system partially really recognized, evaluated, or reproduced my answers.
- Surprisingly human behaviour
- Good recognition of my input, and good response to almost all my counter questions (encountered a bug after a counter question).

- This was good way to cope with my problems. It is very helpful if you have not really dealt with your problem before.
- Less pressure, since the coach is not real and therefore won't judge me and is patient.
- I really liked the structured nature of the conversation and the answers of the bot. The mode of expression was quite realistic.
- Very positive: the rhetoric of and the way he led the conversation. generally a realistic feeling. The whole conversation and the decision making process was well organized. Partially surprisingly perfectly responses on what I said.
- The summary of my problem and the resulting customized questions were very helpful.

##### 7.4.2 Negative Feedback

- The bot did not really get into what i said. The questions were rather general and predictable.
- The decision helper reacted rather predictable. Although generally helpful, it was quite awkward to speak to him, since he sometimes did not reacted appropriately.
- The definition of risk was unclear. Vpino could not really clarify what he meant when asked for what he meant with risk.
- The conversation was too "static", unlike chatting with a human.
- No chance of correcting my input
- At the beginning, i did not really know how to respond or what i was expected to say.
- The coaching took way too long for my problem
- He did not "understand" me
- Resonse times of the system were quite irritating.
- He did't get my sarcasm.
- Did not include any knowledge from external databases (for example buying a TV), which would be very desirable for such a tool.
- The general nature of the conversation :-(
- Vpino did not really get into my problem and it was impossible to change or correct my input.
- He would not give a suggestion, neither did he explain "risk"

## 7.5 Discussion and Conclusions

The results of the user study presented in this chapter verified the findings from the previous study. In general, the system works best for motivated users that are open for new experience with a chat coach. Apart from that, more hypervigilant participants benefit particularly by reduced emotional pressure after receiving a coaching by VPINO. In general, VPINO did not perform too well for rather emotional problems. The study also revealed that there is a difference in how users engage with VPINO. Whereas the structured nature of the conversation was valued by a large number of users, others disliked the structure for being too structured or static. However, a structured approach is the key concept and the actual intention of VPINO. Some users expected the coach to provide a solution for their particular problem or a suggestion for their problem. Anyway, suggesting a particular solution is not the intended behaviour for a computer coach, since professional human coach would not give a concrete recommendation either. VPINO's asynchronous turn-taking mechanism with delayed response time was designed to imitated a human chat partner. Conversation with VPINO was rated positively natural by many users. Other participants found this exact behaviour more irritating than helpful, since they expect a chatbot to respond instantly. Some of the appreciation and especially criticism of VPINO's coaching approach seemed to a matter of taste and user expectations. To conclude, an optimal chat conversation is held when a client is clearly aware of the intentions and methodology of the coach, and are open and motivated to work with the coach.



## CHAPTER 8

# Software Architecture

---

This chapter describes the software architecture and tools developed through the course of this work. It is recommended for interested readers that are familiar with technical terms.

### 8.1 Communication and Software Architecture of VPINO

This section gives a description of the software architecture of the VPINO dialogue system. Figure 8.1 provides an overview on the software components and the communication in VPINO's. The components displayed in the figure are described in the following sections.

#### 8.1.1 User Interface

VPINO is presented to the user via a lightweight web-interface using Bootstrap 2.3.2<sup>1</sup>, as shown in Fig. 1.1. The interface is based on a web-application written in Ruby 1.9.3<sup>2</sup>, using the Sinatra Framework<sup>3</sup>. It was designed to look familiar and similar to common messaging-interfaces as known from common chat programs.

To start a session with VPINO, the user logs on the website. In the background, the browser-client connects to a messaging server using Javascript Faye client, where it subscribes to a dedicated channel (See 8.1.2). The user interface exchanges messages between the user and the subscribed channel.

As an alternative to the web-interface, we provide a XMPP<sup>4</sup>/Jabber client. Furthermore, we have implemented a prototypical interface for spoken conversation, which will

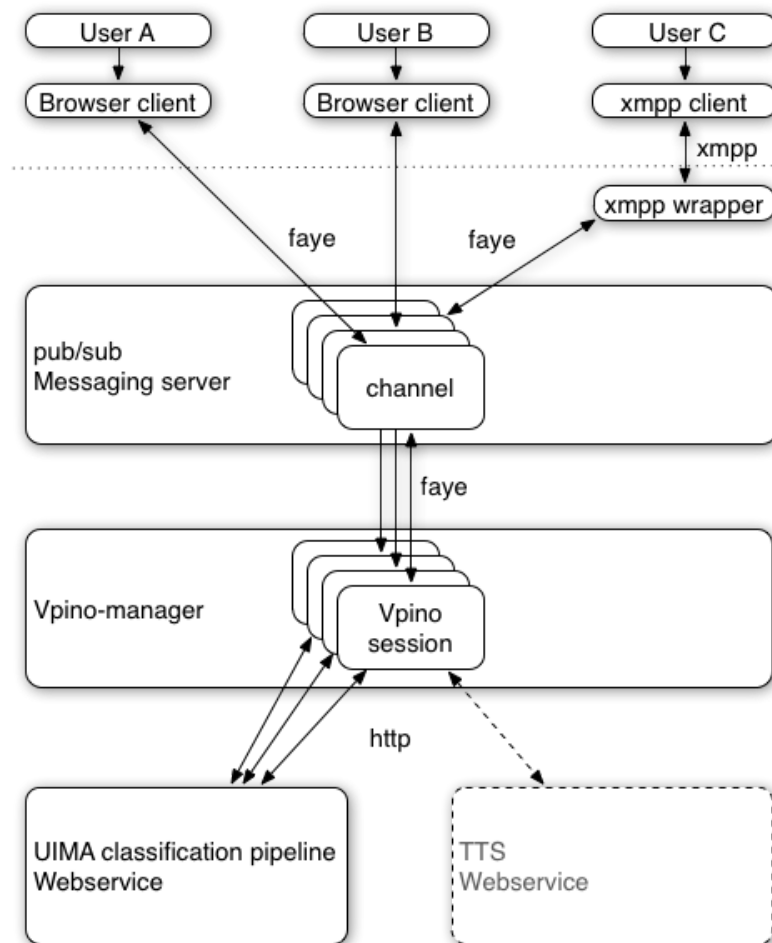
---

<sup>1</sup><http://getbootstrap.com/2.3.2>

<sup>2</sup><http://www.ruby-lang.org>

<sup>3</sup><http://www.sinatrarb.com>

<sup>4</sup><http://www.xmpp.org>



**Figure 8.1:** Overview of the software communication architecture. User clients subscribe to a dedicated channel on a pub/sub server. The VPINO session-manager creates and subscribes VPINO sessions to channels. VPINO sessions access a single instance of the UIMA dialogue act classification component via a http web-service. The optional text-to-speech(TTS) component is also provided as a web-service.

be presented in Sect. 8.1.5. However, for the user studies presented in this thesis, the web-interface was used.

### 8.1.2 Publish/Subscribe Messaging Server

To manage communication between the user interface and the agent, VPINO uses a publish/subscribe (pub/sub) messaging system (Faye<sup>7</sup>). Pub/sub messaging enables asynchronous communication between the interface and VPINO. The communication server accepts a new session request from a user-client and creates a new channel to which the user client subscribes. An instance of VPINO is subscribed by the VPINO session-manager to the newly created channel. Once subscribed, the user-interface and a unique instance of VPINO are able to publish messages and listen to all incoming messages on that channel instantly, without timing restrictions (except network latency). Thereby, timing, turn-taking and response generation are independent from the technical communication component, and can be handled in the particular instance of VPINO.

### 8.1.3 VPINO Session-Manager

VPINO operates as a parallel multi-user system. The session-manager creates and handles VPINO's session instances for each session. Like the user interface, session-manager is a client to the pub/sub-messaging server. The manager listens to the pub/sub-server for new user registrations, creates a new instance of a VPINO session and subscribes this instance to the corresponding communication channel on the Pub/Sub server. Each VPINO instance provides its own message queue, turn-taking manager, structured knowledge memory, sub-dialogue states and conversation history.

### 8.1.4 Dialogue Act Classifier

The dialogue act classification component, as introduced in Sect. 2.3, is provided as a web-service and uses the http protocol for communication. Therefore, it is only instantiated once and used by all VPINO sessions in parallel.

### 8.1.5 Speech-based Communication

VPINO is also capable of spoken communication with the user. However, the speech-based interface is a prototype and was not used in the user studies.

The speech-based interface is an audio output extension of the web-interface. At the moment, there is no stand-alone version for this speech-based coaching yet. In order to use spoken communication with VPINO, the user is required to have suitable speech recognition software installed on the client system. The web-interface basically talks back by

---

<sup>7</sup><http://faye.jcoglan.com>

reading out the text-responses by VPINO. Client requirement for audio output by VPINO is a html5 compatible browser.

We have integrated the open source text-to-speech (TTS) software tool MaryTTS<sup>5</sup> to enable audio output for VPINO. The TTS component provided as a web-service.

VPINO's text-messages, which are exchanged via the pub/sub-server, include html-tags for highlighting and formatting. Therefore, the response texts are normalized and optimized for TTS processing. For each response message generated by VPINO, the text is immediately processed by the TTS component. The TTS component creates a unique audio resource for each new response text. In order to reduce computational effort and space restrictions, audio files for static texts are buffered. The URL<sup>6</sup> of the created audio file is attached to the original message, and is finally sent to the user interface via the pub/sub-server.

On the client side, a javascript tool manages the timing of VPINO's utterances. In order to avoid overlap of audio output from multiple messages, the javascript tool keeps a queue for playing the messages in the right order, only one at a time.

Although we have successfully implemented and tested this feature, it was not used in the user studies due to various reasons. Firstly, the additional processing step from speech to text is a potential source of errors. Badly or false recognized words potentially produces problems with VPINO's dialogue act classification component. Speaking to VPINO takes much longer than holding a chat conversation with the system. Also, the preformulated responses by VPINO are optimized for text-based conversation. For example, reading out a formatted list of pros and cons does not generate any additional value, compared to just displaying the list.

## 8.2 Authoring DEV Tool

The development of a natural language dialogue is a lot of manual work for the author. Regardless of the framework or technology used, natural language content (i.e. system responses), and also underlying domain-related models/concepts have to be authored by a human domain expert. In the course of development of VPINO, we have implemented an authoring tool for structured coaching conversation, called (*DEV-tool*). The DEV-tool is used for the creation of conversation models, i.e. sub-dialogues, organisation of sub-dialogues in a dialogue plan, and include software components for further processing and evaluation of user responses.

---

<sup>5</sup><http://mary.dfki.de/>

<sup>6</sup>unique resource locator

V'Pino Webchat - DEV    Sammler -    Subdialogs -

## 13-Problem\_short

Vielleicht können Sie ja noch ein paar kurze Sätze mehr dazu sagen.

Form    Json

New State:

Del Edge:

New Edge:

text[0]

Die Frage war, ob Sie noch mehr zu Ihrem Entscheidungsproblem schreiben können.

text[1]

Gibt es noch weitere Informationen zu Ihrem Entscheidungsproblem?

NEW TEXT:

**Figure 8.2:** User interface of the web-based sub-dialogue DASM authoring tool for modification of dialogue act sequences, and VPINO's response texts at a particular state of the DASM.

### 8.2.1 Sub-Dialogue Editor

The DEV-tool provides a web-interface to create and edit sub-dialogues. It provides a set of forms to edit structure of the DASM, response formulations, as well as code-scripts for the grounding and evaluation phases. For the creation of sub-dialogues, the DEV-tool provides a set of templates for different DASM categories of questions, for example Yes/No-Questions, Numeric-Questions, or Verification-Questions. A new sub-dialogue is created by defining the initial coaching question and the general type of the DASM from the list of templates. Nevertheless, the templates only contain place-holders that have to be manually filled with responses-texts for that particular sub-dialogue. Furthermore, the DASM (i.e. states and dialogue act transitions) has to be customized after creation from a template.

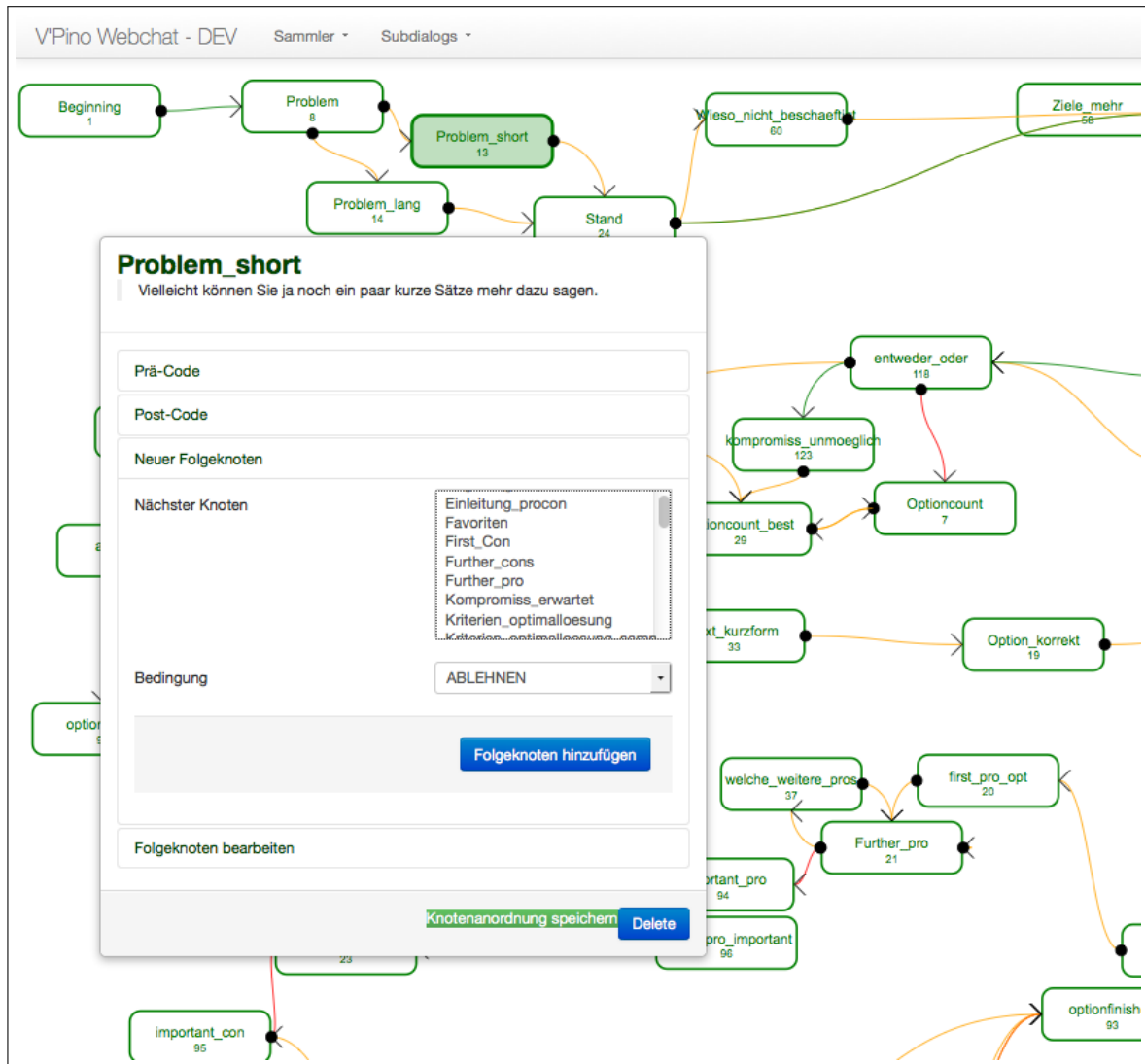
Fig. 8.2 shows a screenshot of the sub-dialogue editor perspective. In view presented in the screenshot, the author edits sub-dialogue identified as *13-Problem\_short*. An overview of the DASM is displayed at the bottom of the screenshot. The header of the view displays information on the current edit perspective. The author is currently editing text for potential system response, for state named *refrage*, changing to the state named *frage*. Sub-dialogues can also be edited and stored in the JSON<sup>7</sup> format. However, editing raw JSON is not recommended due to readability and potential logic inconsistencies in the resulting DASMs.

### 8.2.2 Dialogue-Plan Editor

The DEV-tool is also used to arrange sub-dialogues into a structured dialogue plan. Many sub-dialogues have strict precedence constraints, either because a sub-dialogue requires a specific piece of information in the SKM, or because it is directly related to a preceding question. A screenshot of the dialogue-plan editor is provided in Fig. 8.3. The background of the screenshot shows an overview on all sub-dialogues and the their precedence relations. With the form in the front, the author can define precedence new relations, for example the follow-up sub-dialogue on a user disagreement (*ABLEHNEN*), or changes in the SKM.

---

<sup>7</sup>JSON is a human readable data format based on key-value pairs



**Figure 8.3:** Screenshot of the dialogue-plan editor perspective. The editor is provided as a web-based user interface for authoring sequences of sub-dialogues, pre and post-conditions for the state of the conversation. The screenshot displays an view for a sub-dialogue called “Problem\_short”. In the view presented here, the author defines a follow up sub-dialogue for the particular case when the current sub-dialogue was finally answered with a user disagreement.

## CHAPTER 8. SOFTWARE ARCHITECTURE



## CHAPTER 9

# Discussion and Conclusions

---

The goal of this work was to explore the potential for conversational agents for goal directed coaching conversation. This chapter summarizes the main findings and conclusions presented in the course of this work and discusses open issues and potential future work.

**VPINO** In this thesis we present VPINO, a conversational agent specifically developed for structured coaching conversation. Conversation with VPINO aims to foster reflection by the client by using the technique of Socratic questioning. A coaching conversation with VPINO is constructed by a manually authored set of coaching-question, which are organised in sub-dialogues. VPINO uses a flexible turn-taking strategy in order to imitate human chat behaviour. The basic strategy of the coach is to keep control over the conversation at all times, while using shallow natural language understanding on the level of dialogue acts. VPINO maintains a structured knowledge memory to store and access the information gathered throughout the conversation.

VPINO was developed with the focus on the application scenarios rational decision-making and training transfer coaching. Nevertheless, it can be adapted to other scenarios for structured conversation, where deep semantic understanding is not an essential requirement.

**General effectiveness and usefulness** For the two presented scenarios of rational decision-making and training transfer, we have conducted a series of user studies. Thereby, we could show that VPINO can effectively support humans as a coach. In the training transfer scenario, daily coaching with VPINO successfully improved the communication skills of the users after one week. Furthermore, when compared to a conventional transfer strategy, higher login rates with VPINO indicate that it was more motivating. The training transfer scenario is a good example for the intended area of use for VPINO. The transfer of newly learned communication-skills into everyday work is a highly individual task, ideally sup-

ported by a human coach. However, the use of a human coach is impractical for daily, but relatively short coaching sessions for a large number of users. VPINO, in contrast, can serve as a cheap, highly available, low threshold substitute.

Whereas transfer coaching sessions are rather short and focus on the fixed topic of communication skills, rational decision coaching is a more challenging scenario. VPINO employs a vigilant decision-making strategy, adopted from Benjamin Franklin's Pros and Cons method, in addition with a goal setting approach. The subject of the conversation with VPINO, i.e. the particular decision problem, is freely chosen by the user. In our user studies, VPINO was evaluated as useful and effectively supported many users with making a decision or planning the next steps, even without understanding the details of the users' actual decision problem. Decision coaching with VPINO was particularly successful for participants, that generally have a more hurried and anxious approach on decisions. However, for users with more emotional than rational problems, VPINO's lack of empathy resulted in less human-like conversation, and therefore unsatisfied users. On the other hand, rational decision coaching was never intended for rather emotional than rational problems, anyway.

**User attitude and expectations** Conversation with a computer based coach does not work out equally well for every user. New technology is always challenging the curiosity of humans using that technology. The usefulness of VPINO is influenced by the users' attitude towards using a computer coach. More specifically, participants that are more open to new technology are more likely to profit from coaching with VPINO. Sceptical users, on the contrary, were more likely to produce less human-like and ineffective conversations.

In general, a fundamental requirement for successful conversation is a high degree of motivation and willingness to cooperate from the user. However, this is also a requirement for personal coaching with a human coach.

Transparency is also a key factor for success of a coaching with VPINO. User expectations on computer based coaching are often based on vague notions on the state of the art in technology. Some users expect human-like deep understanding about their problem, others expected VPINO to be a question-answering or information retrieval system. In order to avoid misunderstandings, there is a strong need to manage the users expectations, i.e. give a detailed clarification of the scope of the conversation and the methodology used. Especially, users have to be aware that VPINO will not calculate and suggest a "optimal" solution. Besides an explicit clarification, we consider the extensive use of grounding techniques as a suitable strategy to clarify the intentions and capabilities of VPINO.

**Chat language** In chat-based communication, the use of language largely differs from language in "traditional" contexts, e.g. in text documents, spoken human-to-human conversation. Depending on the age, experience and everyday use of chat-based communication, chat-language varies between humans with respect to grammatical and syntactical

correctness, style, length of responses, and politeness/respect. In general, chat language shows a tendency towards the use of abbreviation and acronyms. Throughout all user-studies conducted in this thesis, the majority of users ignored case sensitive writing. Also, spelling errors were commonly made. Whereas some user responded with small essays on simple yes/no-question, others preferred responding with merely grammatically correct sentences or a bag of keywords. A general observation was that older participants more often responded with longer utterances, and furthermore tend to a more formally correct use of language.

The inconsistent use of language by the users is a challenge for natural language human-computer interaction. In order to adapt a computer coach to different types of users, further work should investigate the detection and categorization of user chat-communication style in an early stage of the conversation.

**Authoring effort** In contrast to data-centric information retrieval systems, professional applications for holding task-oriented conversation still require a large amount of effort from human authors for the development of domain models, matching patterns, and response texts. VPINO does not constitute an exception: The set of rules for VPINO's dialogue act classification, the dialogue act sequence models, sub-dialogues, VPINO's questions, as well as of its potential responses, have to be authored by a human expert.

**Outlook and future work** While this work focused on text-based chat communication, future work may include spoken computer based coaching conversation. Text-to-speech and speech-recognition algorithms have reached a quality that allows their practical application as natural interfaces. Although we have tested a first prototype for a speech-based interface, there are still unsolved speech-related problems in the field of conversation, such as robust turn-taking strategies and avoiding speech overlap.

Furthermore, future work may also investigate the potential of supporting group-decisions, which is a great challenge with respect to turn-taking and organisation of discourse.

In human face-to-face coaching, the physical collocation of the coach and the client allows the coach to make use of tools. In addition to just talking, a coaching session can include the use of flip-charts or, even more simple, pen and paper. A computer based-coach and a human client are virtually collocated with a web-based user interface. Further research should investigate multi-modal approaches that use the flexibility of web-based interfaces to extended purely chat-based conversation with helpful interactive tools. As an example for decision coaching, a multi-modal user interface could display an interactive map of the problem, the options and arguments.

## CHAPTER 9. DISCUSSION AND CONCLUSIONS

# Bibliography

---

- [Aiken and West, 1991] Aiken, L. S. and West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.
- [Aleven and Koedinger, 2002] Aleven, V. A. and Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive science*, 26(2):147–179.
- [Allen and Core, 1997] Allen, J. and Core, M. (1997). Draft of damsl: Dialog act markup in several layers.
- [Allen et al., 1999] Allen, J., Guinn, C. I., and Horvtz, E. (1999). Mixed-initiative interaction. *Intelligent Systems and their Applications, IEEE*, 14(5):14–23.
- [Allwood, 1976] Allwood, J. (1976). *Linguistic communication as action and cooperation*. University of Göteborg. Department of Linguistics.
- [Austin, 1962] Austin, J. (1962). *How to do things with words*. Oxford University Press.
- [Baldwin and Ford, 1988] Baldwin, T. T. and Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel psychology*, 41(1):63–105.
- [Barnett and Ceci, 2002] Barnett, S. M. and Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, 128(4):612.
- [Berg and Szabo, 2005] Berg, I. K. and Szabo, P. (2005). *Brief coaching for lasting solutions*. WW Norton New York.
- [Bickmore et al., 2011] Bickmore, T. W., Schulman, D., and Sidner, C. L. (2011). A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of biomedical informatics*, 44(2):183–197.
- [Brennan, 1998] Brennan, S. E. (1998). The grounding problem in conversations with and through computers. *Social and cognitive approaches to interpersonal communication*, pages 201–225.

## BIBLIOGRAPHY

- [Burke and Hutchins, 2007] Burke, L. A. and Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human resource development review*, 6(3):263–296.
- [Cahn and Brennan, 1999] Cahn, J. E. and Brennan, S. E. (1999). A psychological model of grounding and repair in dialog. In *Proc. Fall 1999 AAAI Symposium on Psychological Models of Communication in Collaborative Systems*.
- [Carey et al., 2011] Carey, W., Philippon, D. J., and Cummings, G. G. (2011). Coaching models for leadership development: An integrative review. *Journal of Leadership Studies*, 5(1):51–69.
- [Carletta et al., 1997] Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J. C., and Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational linguistics*, 23(1):13–31.
- [Chiticariu et al., 2013] Chiticariu, L., Li, Y., and Reiss, F. R. (2013). Rule-based information extraction is dead! long live rule-based information extraction systems! In *EMNLP*, pages 827–832.
- [Clark and Brennan, 1991] Clark, H. H. and Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149.
- [Clark and Schaefer, 1989] Clark, H. H. and Schaefer, E. F. (1989). Contributing to discourse. *Cognitive science*, 13(2):259–294.
- [Clemen, 1996] Clemen, R. (1996). *Making Hard Decisions: An Introduction to Decision Analysis*. ed. Boston: Duxbury Press.
- [Clutterbuck and Hussain, 2010] Clutterbuck, D. and Hussain, Z. (2010). *Virtual coach, virtual mentor*. IAP.
- [Colby, 1981] Colby, K. M. (1981). Modeling a paranoid mind. *Behavioral and Brain Sciences*, 4(04):515–534.
- [Core and Allen, 1997] Core, M. and Allen, J. (1997). Coding dialogs with the damsl annotation scheme. In *AAAI fall symposium on communicative action in humans and machines*, pages 28–35. Boston, MA.
- [Costa and McCrae, 1992] Costa, P. T. and McCrae, R. R. (1992). Four ways five factors are basic. *Personality and individual differences*, 13(6):653–665.
- [Cramton, 2001] Cramton, C. D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization science*, 12(3):346–371.

- [de Castilho and Gurevych, 2009] de Castilho, R. E. and Gurevych, I. (2009). Dkpro-ugd: A flexible data-cleansing approach to processing user-generated discourse. In *Online-proceedings of the First French-speaking meeting around the framework Apache UIMA*, LINA CNRS UMR 6241 - University of Nantes, Nantes, France.
- [De Shazer and Dolan, 2012] De Shazer, S. and Dolan, Y. (2012). *More than miracles: The state of the art of solution-focused brief therapy*. Routledge.
- [Devaraj et al., 2008] Devaraj, S., Easley, R. F., and Crant, J. M. (2008). How does personal-ity matter? relating the five-factor model to technology acceptance and use. *Information Systems Research*, 19(1):93–105.
- [Duncan, 1972] Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2):283.
- [Ericsson et al., 1993] Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3):363.
- [Ericsson and Simon, 1980] Ericsson, K. A. and Simon, H. A. (1980). Verbal reports as data. *Psychological review*, 87(3):215.
- [Ferrucci et al., 2010] Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyan-pur, A. A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., et al. (2010). Building watson: An overview of the deepqa project. *AI magazine*, 31(3):59–79.
- [Ferrucci and Lally, 2004] Ferrucci, D. and Lally, A. (2004). Uima: an architectural ap-proach to unstructured information processing in the corporate research environment. *Natural Language Engineering*, 10(3-4):327–348.
- [Figueira et al., 2005] Figueira, J., Greco, S., and Ehrgott, M. (2005). *Multiple criteria decision analysis: state of the art surveys*, volume 78. Springer.
- [Fournies, 1987] Fournies, F. F. (1987). *Coaching for improved work performance*. Liberty House.
- [Geißler and Metz, 2012] Geißler, H. and Metz, M. (2012). *E-Coaching und Online-Beratung: Formate, Konzepte, Diskussionen*. Springer-Verlag.
- [Gigerenzer and Goldstein, 1996] Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4):650.
- [Godfrey et al., 1992] Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, volume 1, pages 517–520. IEEE.

## BIBLIOGRAPHY

- [Graesser et al., 2005] Graesser, A. C., Chipman, P., Haynes, B. C., and Olney, A. (2005). Autotutor: An intelligent tutoring system with mixed-initiative dialogue. *Education, IEEE Transactions on*, 48(4):612–618.
- [Graesser et al., 2004] Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., and Louwerse, M. M. (2004). Autotutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, 36(2):180–192.
- [Graesser et al., 1999] Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., Group, T. R., et al. (1999). Autotutor: A simulation of a human tutor. *Cognitive Systems Research*, 1(1):35–51.
- [Grant, 2003] Grant, A. M. (2003). The impact of life coaching on goal attainment, metacognition and mental health. *Social Behavior and Personality: an international journal*, 31(3):253–263.
- [Grant et al., 2010] Grant, A. M., Passmore, J., Cavanagh, M. J., Parker, H. M., et al. (2010). The state of play in coaching today: A comprehensive review of the field. *International Review of Industrial and Organizational Psychology*, 25(1):125–167.
- [Grice, 1970] Grice, H. P. (1970). *Logic and conversation*.
- [Grosz et al., 1977] Grosz, B. J., Hendrix, G. G., and Robinson, A. E. (1977). Using process knowledge in understanding task-oriented dialogs. In *IJCAI*, page 90.
- [Haccoun, 1997] Haccoun, R. R. (1997). Transfer and retention: Let’s do both and avoid dilemmas. *Applied Psychology*, 46(4):340–344.
- [Hamlin et al., 2008] Hamlin, R. G., Ellinger, A. D., and Beattie, R. S. (2008). The emergent ‘coaching industry’: a wake-up call for hrd professionals. *Human Resource Development International*, 11(3):287–305.
- [Han et al., 2013] Han, L., Kashyap, A., Finin, T., Mayfield, J., and Weese, J. (2013). Umbc ebiquity-core: Semantic textual similarity systems. *Atlanta, Georgia, USA*, 44.
- [Hanington and Martin, 2012] Hanington, B. and Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers.
- [Herold et al., 2002] Herold, D. M., Davis, W., Fedor, D. B., and Parsons, C. K. (2002). Dispositional influences on transfer of learning in multistage training programs. *Personnel Psychology*, 55(4):851–869.
- [Hutchens and Alder, 1998] Hutchens, J. L. and Alder, M. D. (1998). Introducing megal. In *Proceedings of the Joint Conferences on New Methods in Language Processing and*



- Computational Natural Language Learning*, pages 271–274. Association for Computational Linguistics.
- [IBM, 2015] IBM (2015). Cognitive computing. Online, Retrieved August 1, 2015, <http://www.research.ibm.com/cognitive-computing>.
- [ICF, 2015] ICF (2015). What is professional coaching? Online, Retrieved August 1, 2015, from "<http://www.coachfederation.org/need/landing.cfm?ItemNumber=978&navItemNumber=567>".
- [Ives, 2008] Ives, Y. (2008). What is ‘coaching’? an exploration of conflicting paradigms. *International Journal of Evidence Based Coaching and Mentoring*, 6(2):100–113.
- [Janis, 1982] Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes*. Houghton Mifflin Boston.
- [Janis and Mann, 1977] Janis, I. L. and Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press.
- [Keith et al., 2010] Keith, N., Richter, T., and Naumann, J. (2010). Active/exploratory training promotes transfer even in learners with low motivation and cognitive ability. *Applied Psychology*, 59(1):97–123.
- [Kim et al., 2010] Kim, S. N., Cavedon, L., and Baldwin, T. (2010). Classifying dialogue acts in one-on-one live chats. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 862–871. Association for Computational Linguistics.
- [Kluegl et al., 2015] Kluegl, P., Toepfer, M., Beck, P.-D., Fette, G., and Puppe, F. (2015). Uima ruta: Rapid development of rule-based information extraction applications. *Natural Language Engineering*, FirstView:1–40.
- [L’Abbate et al., 2005] L’Abbate, M., Thiel, U., and Kamps, T. (2005). Can proactive behavior turn chatterbots into conversational agents? In *Intelligent Agent Technology, IEEE/WIC/ACM International Conference on*, pages 173–179. IEEE.
- [Landauer and Dumais, 1997] Landauer, T. K. and Dumais, S. T. (1997). A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211.
- [Latorre-Navarro and Harris, 2014] Latorre-Navarro, E. M. and Harris, J. G. (2014). A natural language conversational system for online academic advising. In *The Twenty-Seventh International Flairs Conference*.
- [Lin, 2004] Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop*, volume 8.

## BIBLIOGRAPHY

- [Luhmann, 1984] Luhmann, N. (1984). *Soziale systeme*. Suhrkamp Frankfurt am Main.
- [Mann et al., 1997] Mann, L., Burnett, P., Radford, M., and Ford, S. (1997). The melbourne decision making questionnaire: An instrument for measuring patterns for coping with decisional conflict. *Journal of Behavioral Decision Making*, 10(1):1–19.
- [Marineau et al., 2000] Marineau, J., Wiemer-Hastings, P., Harter, D., Olde, B., Chipman, P., Karnavat, A., Pomeroy, V., Graesser, A., and the TRG (2000). Classification of speech acts in tutorial dialog. In *Proceedings of the workshop on modeling human teaching tactics and strategies at the Intelligent Tutoring Systems 2000 conference*, pages 65–71.
- [Mateas and Stern, 2003] Mateas, M. and Stern, A. (2003). Façade: An experiment in building a fully-realized interactive drama. In *Game Developers Conference*, pages 4–8.
- [Mäurer et al., 2013] Mäurer, D., Bruns, A., and Weihe, K. (2013). Be open to computer based coaching. In *Proceedings of SemDIAL 13*, pages 102–111.
- [Mäurer and Weihe, 2015] Mäurer, D. and Weihe, K. (2015). Benjamin Franklin’s decision method is acceptable and helpful with a conversational agent. In *Intelligent Interactive Multimedia Systems and Services*, pages 109–120. Springer.
- [McGuire, 1960] McGuire, W. J. (1960). Cognitive consistency and attitude change. *The Journal of Abnormal and Social Psychology*, 60(3):345.
- [Mollá and Vicedo, 2007] Mollá, D. and Vicedo, J. L. (2007). Question answering in restricted domains: An overview. *Computational Linguistics*, 33(1):41–61.
- [Morbini et al., 2014] Morbini, F., DeVault, D., Sagae, K., Gerten, J., Nazarian, A., and Traum, D. (2014). Flores: A forward looking, reward seeking, dialogue manager. In *Natural Interaction with Robots, Knowbots and Smartphones*, pages 313–325. Springer.
- [Neenan and Palmer, 2001] Neenan, M. and Palmer, S. (2001). Cognitive behavioural coaching<sup>1</sup>.
- [Neumann and Morgenstern, 1947] Neumann, L. J. and Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton University Press Princeton, NJ.
- [Nickols, 2015] Nickols, F. (2015). How to make decisions like Benjamin Franklin. Online, Retrieved August 1, 2015, from "[www.smartdraw.com/mind-map/making-decisions-ben-franklin.htm](http://www.smartdraw.com/mind-map/making-decisions-ben-franklin.htm)".
- [Olivero et al., 1997] Olivero, G., Bane, K. D., and Kopelman, R. E. (1997). Executive coaching as a transfer of training tool: Effects on productivity in a public agency. *Public personnel management*, 26:461–470.

- [O'Shea et al., 2012] O'Shea, J., Bandar, Z., and Crockett, K. (2012). A multi-classifier approach to dialogue act classification using function words. In *Transactions on Computational Collective Intelligence VII*, pages 119–143. Springer.
- [Paek and Horvitz, 1999] Paek, T. and Horvitz, E. (1999). Uncertainty, utility, and misunderstanding: A decision-theoretic perspective on grounding in conversational systems. In *AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems, Cape Cod, MA*.
- [Paul and Binker, 1990] Paul, R. W. and Binker, A. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world*. ERIC.
- [Popescu-Belis, 2003] Popescu-Belis, A. (2003). Dialogue act tagsets for meeting understanding: an abstraction based on the damsl, switchboard and icsi-mr tagsets. *Report IM2. MDM-09*.
- [Radatz, 2003] Radatz, S. (2003). Beratung ohne ratschlag. *Systemisches Coaching für Führungskräfte und BeraterInnen. Ein Praxishandbuch mit den Grundlagen systemisch-konstruktivistischen Denkens, Fragetechniken, Coachingkonzepten, Rahmenbedingungen von Coachings: Spielregeln, Haltung, Dauer*.
- [Rammstedt and John, 2007] Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, 41(1):203–212.
- [Redshaw, 2000] Redshaw, B. (2000). Do we really understand coaching? how can we make it work better? *Industrial and Commercial training*, 32(3):106–109.
- [Reithinger and Klesen, 1997] Reithinger, N. and Klesen, M. (1997). Dialogue act classification using language models. In *EuroSpeech*. Citeseer.
- [Rizzo et al., 2011] Rizzo, A., Sagae, K., Forbell, E., Kim, J., Lange, B., Buckwalter, J., Williams, J., Parsons, T., Kenny, P., Traum, D., et al. (2011). Simcoach: an intelligent virtual human system for providing healthcare information and support. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, volume 2011. NTSA.
- [Sacks et al., 1974] Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735.
- [Sadek, 1991] Sadek, M. D. (1991). Dialogue acts are rational plans. In *The Structure of Multimodal Dialogue; Second VENACO Workshop*.
- [Saygin et al., 2003] Saygin, A. P., Cicekli, I., and Akman, V. (2003). Turing test: 50 years later. In *The Turing Test*, pages 23–78. Springer.

## BIBLIOGRAPHY

- [Schegloff, 1968] Schegloff, E. A. (1968). Sequencing in conversational openings<sup>1</sup>. *American anthropologist*, 70(6):1075–1095.
- [Schegloff, 1990] Schegloff, E. A. (1990). On the organization of sequences as a source of “coherence” in talk-in-interaction. *Conversational organization and its development*, 38:51–77.
- [Schmid, 1994] Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing*, volume 12, pages 44–49. Citeseer.
- [Schraw and Moshman, 1995] Schraw, G. and Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, 7(4):351–371.
- [Schulman et al., 2011] Schulman, D., Bickmore, T. W., and Sidner, C. L. (2011). An intelligent conversational agent for promoting long-term health behavior change using motivational interviewing. In *AAAI Spring Symposium: AI and Health Communication*.
- [Schwartz and Kliban, 2004] Schwartz, B. and Kliban, K. (2004). The paradox of choice: Why more is less. Ecco New York.
- [Searle, 1969] Searle, J. (1969). *Speech Acts*. Cambridge University Press, Cambridge, Mass.
- [Searle, 1980] Searle, J. (1980). Minds brains and machines. *The Behavioural and Brain Sciences*, 3(3):417–457.
- [Shawar and Atwell, 2007] Shawar, B. A. and Atwell, E. (2007). Chatbots: are they really useful? In *LDV Forum*, volume 22, pages 29–49.
- [Simon, 1979] Simon, H. A. (1979). Rational decision making in business organizations. *The American economic review*, pages 493–513.
- [Sparks, 1844] Sparks, J. (1844). *The Works of Benjamin Franklin*.
- [Stalnaker, 1978] Stalnaker, R. (1978). Assertion.
- [Stewart et al., 2008] Stewart, L. J., Palmer, S., Wilkin, H., and Kerrin, M. (2008). The influence of character: Does personality impact coaching success. *International journal of evidence based coaching and mentoring*, 6(1):32–42.
- [Stolcke et al., 2000] Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. V., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.

- [Traum et al., 1999] Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999). A model of dialogue moves and information state revision. Technical report, Citeseer.
- [Traum, 1999] Traum, D. R. (1999). Computational models of grounding in collaborative systems. In *Psychological Models of Communication in Collaborative Systems-Papers from the AAAI Fall Symposium*, pages 124–131.
- [Traum, 2000] Traum, D. R. (2000). 20 questions on dialogue act taxonomies. *Journal of semantics*, 17(1):7–30.
- [Traum and Hinkelman, 1992] Traum, D. R. and Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational intelligence*, 8(3):575–599.
- [Turing, 1950] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.
- [Von Neumann and Morgenstern, 1944] Von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton university press.
- [Wallace, 2009] Wallace, R. S. (2009). *The anatomy of ALICE*, chapter 13, pages 181–210. Springer.
- [Weizenbaum, 1966] Weizenbaum, J. (1966). Eliza - a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45.
- [Wexley and Baldwin, 1986] Wexley, K. N. and Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal*, 29(3):503–520.
- [Whitmore, 2009] Whitmore, J. (2009). Coaching for performance: Growing human potential and purpose: The principles and practices of coaching and leadership. Boston, MA: Nicholas Brealey.
- [Yankelovich et al., 1995] Yankelovich, N., Levow, G.-A., and Marx, M. (1995). Designing speechacts: Issues in speech user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 369–376. ACM Press/Addison-Wesley Publishing Co.
- [Zhou and Oldham, 2001] Zhou, J. and Oldham, G. R. (2001). Enhancing creative performance: Effects of expected developmental assessment strategies and creative personality. *The Journal of Creative Behavior*, 35(3):151–167.

## BIBLIOGRAPHY

## List of Figures

---

1.1	VPINO web-based user interface . . . . .	7
2.1	VPINO dialogue act classification pipeline . . . . .	26
2.2	Translated example transcript for sub-dialogue from user study . . . . .	33
2.3	Original German example transcript for sub-dialogue from user study . . . . .	34
2.4	Example dialogue act sequence model (DASM), real example taken from the user study . . . . .	35
2.5	Schematic overview of a sub-dialogue . . . . .	38
2.6	State chart for VPINO's turn-taking and user communication protocol . . . . .	40
2.7	Screenshot of the user interface showing a visual hint for delayed system responses . . . . .	41
4.1	User interface of STORI communication training tool. Example critical communication situation at work. . . . .	50
4.2	Transcript of a training transfer session taken from the user study . . . . .	53
4.3	Diagram for the user study design for training transfer coaching. . . . .	54
4.4	Gain in Communication Skills depending on Openness and Transfer Strategy	58
5.1	Example decision making spreadsheet . . . . .	65
5.2	Overview of conversation organization and SKM for rational decision coaching . . . . .	72
6.1	Influence of attitude and cooperation on evaluation results . . . . .	84
6.2	Influence of hypervigilant decision-making approach on evaluation results	85
7.1	Mean combined results for attitude, spwc and hypervigilance . . . . .	94
7.2	User attitude towards using a chat coach . . . . .	95

## LIST OF FIGURES

8.1	Vpino software communication architecture . . . . .	100
8.2	User interface of the web-based sub-dialogue DASM authoring . . . . .	103
8.3	The authoring editor tool for dialogue plans . . . . .	105
B.1	Instructions for the communication training tool . . . . .	127
B.2	Goal setting for the conventional training transfer strategy . . . . .	128
B.3	Instructions for training transfer . . . . .	129
B.4	Conventional training transfer online form-sheet . . . . .	130



## List of Tables

---

2.1	Frequencies of dialogue acts in SWBD-DAMSL corpus . . . . .	21
2.2	Set of dialogue acts used with VPINO . . . . .	24
4.1	Means, Standard Deviations and Pearson Correlations for measures in the user study . . . . .	56
4.2	Test of Moderation Transfer Strategy x Openness on Communication Skills.	57
5.1	Characteristics of vigilant vs. hypervigilant decision-making . . . . .	66
5.2	Option characteristics . . . . .	75
6.1	Evaluation question results . . . . .	82
6.2	Pearson correlations for evaluation measures . . . . .	83
7.1	Evaluation question results for user study 2 . . . . .	92
7.2	Pearson correlations for evaluation measures . . . . .	93

## LIST OF TABLES

## APPENDIX A

# Prestudy

---

The appendix contains additional material that was not included in the main part due to space restrictions. The materials that were used for the user studies are presented in their original form in German language.

### A.1 Systemic Coaching Questions

The list of systemic coaching questions in German language, as used in the coaching sessions with VPINO in the pre-study.

- Gibt es denn Grenzen oder Tabu-Zonen, über die Sie nicht detailliert sprechen möchten?
- Bitte benennen Sie doch mal konkret die Themengebiete, die für Sie tabu sind.
- Bei welchem Problem kann ich ihnen denn versuchen zu helfen?
- Wenn Sie es in einem Satz auf den Punkt bringen müssten, worin besteht dabei der genaue Konflikt?
- Sind möglicherweise noch andere Personen an der Situation beteiligt?
- Welche Ursachen könnte denn Ihr Problem haben?
- Wenn Sie an Ihre aktuelle Situation denken, wo befinden Sie sich dann auf einer Skala von 0 bis 10? (0 -> Die Situation ist ganz fürchterlich, 10 -> die Situation ist optimal)
- Was wäre denn Ihr Wunschwert?
- Was möchten Sie gerne in dieser Sitzung erreichen?
- Gibt es eine Person, die Sie hierbei besonders gut unterstützen kann
- Wie würde ein Außenstehender, der Sie nicht kennt, Ihre Situation einschätzen?

## APPENDIX A. PRESTUDY

- Was müssten Sie aktiv tun, um das Problem zu behalten oder gar zu verschlimmern?
- Was würde Sie in Hinblick auf Ihre Situation noch unglücklicher machen?
- Haben Sie eine vergleichbare Situation schon einmal erlebt?
- Gibt es Menschen, die Sie unterstützen können?
- Was gibt Ihnen zurzeit Kraft dieses Problem zu lösen?
- Beschreiben Sie doch mal ihr eigentliches Ziel bezogen auf ihr Problem.
- Ist das Ziel in Ihren eigenen Handlungsmöglichkeiten realistisch?
- Was möchten Sie in 6 Monaten erreicht haben?
- Woran werden Sie merken, dass Sie Ihr Ziel erreicht haben?
- Wer in Ihrem Umfeld müsste sich wie verhalten, damit Sie Ihr Ziel erreichen?
- Wer in Ihrem Umfeld würde mehr, wer würde gar nicht von der Lösung profitieren?
- Welche Kriterien müsste die Lösung erfüllen, damit sie für Sie eine gute Lösung darstellt?
- Welche Handlungsmöglichkeiten haben Sie um Ihr Ziel zu erreichen? Bitte nennen Sie ein paar Stichpunkte.
- Bitte beschreiben Sie jetzt eine Handlungsoption genauer.
- Angenommen, Sie hätten bereits den Schlüssel zur Lösung Ihres Problems gefunden, was wäre dann anders?
- Nehmen wir mal an, nachdem wir hier fertig sind, gehen Sie zurück in Ihren Alltag. Sie sehen zum Beispiel noch etwas fern oder tun was auch immer Sie tun. Dann gehen Sie ins Bett und schlafen ein. Während Sie schlafen passiert ein Wunder und Ihre Probleme, die Sie heute mit in diese Sitzung gebracht haben, sind weg - einfach so! Aber da das alles passierte, während Sie schliefen, bemerken Sie gar nicht, dass dieses Wunder passiert ist. Wenn Sie jetzt morgens aufwachen, woran würden Sie als erstes merken, dass dieses Wunder geschehen ist?
- Was werden Sie ab morgen tun, um einen konkreten Schritt in Richtung Lösung zu gehen?

## APPENDIX B

# Training Transfer

---

## B.1 Communication Training Tool

2013-05-27 14:45:21 | Deutsch ▼

# Herzlich Willkommen zu SToRI

- dem Online-Kommunikationstraining der TU Darmstadt



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

SToRI steht für **Simulation-based Training on Real-life Incidents** und basiert auf **kritischen Situationen**. Die Rede ist hier von Situationen, auf die man nicht vorbereitet ist. Oft fühlt man sich dabei „überfahren“, wird nervös und versucht, sich schnell eine schlagfertige Reaktion auszudenken, um die Situation bestmöglich zu bewältigen. Leider gelingt es uns aber gerade in kritischen Kommunikationssituationen oft nicht, **schlagfertig zu reagieren**. Tatsächlich fällt uns eine gute Antwort häufig erst später ein.

SToRI bereitet Sie gezielt auf derartige Situationen im Berufsalltag vor. Dafür schildern wir Ihnen verschiedene kritische Situationen, auf die Sie verschieden gut reagieren können. In Abhängigkeit davon, für welche Reaktionsmöglichkeit Sie sich entscheiden, erscheint die nächste Situation. Wir versuchen also, **echte Gesprächsabläufe zu simulieren**. SToRI gibt Ihnen dabei **Rückmeldung** darüber, wie gut Ihre Entscheidung war und erlaubt Ihnen so, sich zu verbessern. Zusätzlich können und sollen Sie die Situation auch noch einmal aus der **Perspektive Ihres Gegenübers** erleben, denn je besser Sie sich in andere hineinversetzen können, desto besser werden Sie kommunizieren. Probieren Sie es aus!

Die **Ergebnisse des Trainings werden wissenschaftlich ausgewertet**. Diese Auswertung erfolgt jedoch **vollständig anonym**, d.h. Ihre persönlichen Daten (Name, E-Mailadresse) werden nicht mit den Trainingsergebnissen verknüpft. Um Ihren Trainingsfortschritt dennoch speichern zu können, verwenden wir einen persönlichen Code, der ebenfalls nicht mit Ihren persönlichen Daten, sondern nur mit Ihren Trainingsergebnissen verknüpft ist.

**Wählen Sie bitte einen Avatar um zu beginnen:**



Figure B.1: Instructions for the communication training tool.

## B.2 Conventional Transfer Strategy

### Willkommen liebe Teilnehmerin, lieber Teilnehmer!

Ich bitte Sie, sich nachfolgend ein konkretes Ziel zu setzen, welches zur Verbesserung Ihrer Kommunikationsfähigkeiten beitragen soll.

Bei der Formulierung Ihres Ziels beachten Sie bitte Folgendes:

1. Suchen Sie sich irgendeine konkrete Kommunikationsfähigkeit aus, welche Sie verbessern möchten.
2. Definieren Sie, auf welchem Weg, Sie diese Kommunikationsfähigkeit trainieren können/wollen. (Bedenken Sie hierbei, dass es ausreichend Möglichkeiten geben soll, in denen die Zielerreichung wahrscheinlich ist.)
3. Wie können Sie überprüfen, ob Sie dieses Ziel erreicht haben? Wie würden Sie den Erfolg messen, anhand welcher Indizien/Kriterien?
4. Formulieren Sie Ihr Ziel in der ersten Person „Ich“ und vermeiden Sie vage, allgemeine Angaben.

Im Verlauf der nächsten Woche soll dieses Ziel erreicht werden. Damit ist die Zielerreichung auf eine Woche befristet.

**Eine Beispiel-Zielformulierung wäre:**

In der nächsten Woche möchte ich meine Argumentationsstärke verbessern, indem ich meine Gegenargumentationen immer konkret an die Aussagen meines Gesprächspartners anknüpfe. Ich erkenne meinen Erfolg daran, dass keiner der Gesprächspartner die Stimme hebt, um sich zu beweisen.

**Formulieren Sie bitte hier Ihr Ziel:**

„In der nächsten Woche möchte ich  verbessern,

indem ich .

Ich erkenne meinen Erfolg daran, dass .“

Abschicken

**Figure B.2:** Goal setting for the conventional training transfer strategy.

---

## Liebe Teilnehmerin, lieber Teilnehmer,

Sie haben kürzlich den ersten/zweiten/dritten Teil des Trainings abgeschlossen. Dabei haben Sie gelernt, in kritischen Situationen erfolgreich zu kommunizieren.

Gerade in der ersten Zeit ist es wichtig, dass Sie Ihre neu erworbenen Fähigkeiten **bewusst in Ihrem Arbeitsalltag einsetzen und üben**. Nur so können nachhaltige Trainingseffekte erzielt werden.

Versuchen Sie deshalb, sich ein **konkretes Ziel für die kommende Arbeitswoche** zu setzen und damit festzulegen, wie Sie Ihre neu erworbenen Kommunikationsfähigkeiten in Ihrem Arbeitsalltag anwenden möchten.

Konzentrieren Sie sich dabei am besten auf eine einzige Kommunikationsfähigkeit und versuchen Sie täglich an diesem Ziel zu arbeiten.

**Ein Lerntagebuch wird Sie bei der Zielerreichung unterstützen.** Das tägliche Ausfüllen des Lerntagebuchs wird nur etwa **5 Minuten** in Anspruch nehmen. Diese 5 Minuten werden aber einen erheblichen Unterschied in Ihrem Trainingserfolg bewirken!

Setzen Sie sich dazu am besten eine **bestimmte Tageszeit** fest, zu der Sie Ihren Zielfortschritt im Lerntagebuch eintragen werden.

Viel Erfolg

[weiter](#)

**Figure B.3:** Screenshot of the instructions for training transfer with online form sheet.

Transfer Tagebuch

Wie leicht ist es Ihnen gestern gefallen, an Ihrem Wochenziel zu arbeiten?

Wie gut ist es Ihnen gestern gelungen, Ihr Wochenziel umzusetzen? Angabe in Prozent (0% = gar nicht gut, 100% = sehr gut)

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Was hat gestern bei der Umsetzung Ihres Wochenziels gut geklappt?

Was hat gestern bei der Umsetzung Ihres Wochenziels nicht so gut geklappt?

Was würden Sie morgen bei der Arbeit an Ihrem Wochenziel vielleicht anders machen wollen?

Wie stark wirkt sich die Arbeit an Ihrem Wochenziel auf Ihre Kommunikationsfähigkeiten aus?

sehr stark	stark	teils teils	kaum	gar nicht
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wie zufrieden sind Sie mit der Umsetzung Ihres Wochenziels insgesamt?

sehr zufrieden	eher zufrieden	teils teils	eher unzufrieden	sehr unzufrieden
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Was hat Sie gestern davon abgehalten, mehr zu erreichen?

Was könnte Ihnen morgen dabei helfen, Ihr Ziel vollständig zu erreichen?

Weiter

**Figure B.4:** Screenshot of the conventional training transfer strategy with an online form sheet as used in the user study.



## APPENDIX C

# Decision Making

---

## C.1 Benjamin Franklin's Letter to Joseph Priestly<sup>1</sup>

To Joseph Priestley

London, September 19, 1772

Dear Sir,

In the Affair of so much Importance to you, wherein you ask my Advice, I cannot for want of sufficient Premises, advise you what to determine, but if you please I will tell you how.

When these difficult Cases occur, they are difficult chiefly because while we have them under Consideration all the Reasons pro and con are not present to the Mind at the same time; but sometimes one Set present themselves, and at other times another, the first being out of Sight. Hence the various Purposes or Inclinations that alternately prevail, and the Uncertainty that perplexes us.

To get over this, my Way is, to divide half a Sheet of Paper by a Line into two Columns, writing over the one Pro, and over the other Con. Then during three or four Days Consideration I put down under the different Heads short Hints of the different Motives that at different Times occur to me for or against the Measure. When I have thus got them all together in one View, I endeavour to estimate their respective Weights; and where I find two, one on each side, that seem equal, I strike them both out: If I find a Reason pro equal to some two Reasons con, I strike out the three. If I judge some two Reasons con equal to some three Reasons pro, I strike out the five; and thus proceeding I find at length where the Ballance lies; and if after a Day or two of farther Consideration nothing new that is of Importance occurs on either side, I come to a Determination accordingly.

And tho' the Weight of Reasons cannot be taken with the Precision of Algebraic Quantities, yet when each is thus considered separately and comparatively, and the whole lies before me, I think I can judge better, and am less likely to take a rash Step; and in fact I have found great Advantage from this kind of Equation, in what may be called Moral or Prudential Algebra.

Wishing sincerely that you may determine for the best, I am ever, my dear Friend,

Yours most affectionately

B. Franklin

---

<sup>1</sup>As published in "The Works of Benjamin Franklin" by [Sparks, 1844]

## C.2 Example Decision Support Conversation with VPINO

An sample conversation from the user study on decision coaching.

VPINO: Hallo.

SIE: Hi!

VPINO: Schön dass Sie da sind!

VPINO: Ich bin Ihr computergesteuerter Entscheidungshelfer. Wollen wir loslegen?

SIE: ja gerne

VPINO: Okay, dann mal los.

VPINO: Formulieren Sie doch mal Ihr Entscheidungsproblem.

SIE: Ich möchte nach meinem studium einen job im ausland. ich weiß allerdings nicht wo und ob das das richtige für mich ist...

VPINO: Vielleicht können Sie ja noch ein paar kurze Sätze mehr dazu sagen.

SIE: nein.

VPINO: Na gut.

VPINO: Haben Sie sich vor unserem Gespräch schon intensiv mit Ihrem Problem beschäftigt?

SIE: jein

VPINO: Also haben Sie sich schon vorher mit dem Problem beschäftigt oder nicht?

SIE: nich so richtig

VPINO: Schade...

VPINO: Wieso nicht?

SIE: weil ich mir einfach die Zeit nicht genommen habe

VPINO: Na wenn das so ist.

VPINO: Mal abgesehen davon, dass Sie eine Entscheidung treffen möchten...

VPINO: Was sind denn Ihre persönlichen Ziele, die Sie im Bezug auf Ihr Entscheidungsproblem erreichen möchten?

SIE: wie meinen sie das?

VPINO: Würde es helfen, wenn ich einfach mal anders frage?

SIE: ja

VPINO: Was sind denn nun Ihre kurz-, mittel- und langfristigen Ziele die Sie erreichen wollen?

SIE: also kurzfristig würde ich gerne natürlich einen gutbezahlten job haben

SIE: mittelfristig könnte es auch eine familie oder so sein

VPINO: Okay.

VPINO: Haben Sie vielleicht auch noch langfristige Ziele?

SIE: ne denk mal nicht

VPINO: Denken Sie, dass es eine realistische Optimallösung gibt, mit der Sie alle Ziele erreichen können?

SIE: ja das hoffe ich doch!!!

VPINO: Sehr gut.

VPINO: Und handelt es sich bei Ihrem Problem um eine typische Entweder-Oder Entscheidung?

SIE: hm nicht wirklich

VPINO: Dann sammeln wir doch mal Ihre Alternative.

VPINO: Und was würden Sie sagen, wie viele realistische Optionen sehen Sie im Moment?

SIE: ich würde sagen so ca 3

VPINO: Sie sehen also im Moment drei mögliche Optionen.

VPINO: Ist das korrekt?

SIE: ja im moment schon

VPINO: Bevor wir jetzt auf die einzelnen Lösungsmöglichkeiten eingehen, sammeln wir erst einmal Ihre Optionen.

VPINO: Dann legen Sie mal los und konkretisieren bitte ihre erste von drei Alternativen.

VPINO: Wie würden Sie diese Option in ein paar kurzen Sätzen beschreiben?

SIE: Ich will eigentlich für eine weile ins ausland. am liebsten ins englischsprachige ausland.

VPINO: Wie würde Sie diese Alternative abkürzen?

SIE: wieso?

VPINO: Damit wir im weiteren Verlauf die Option schnell beim Namen nennen können.

VPINO: Also? Welche konkrete Abkürzung würde Sie denn dieser Option geben, in einem Wort, geben wollen?

SIE: london

VPINO: Bevor wir mit den einzelnen Optionen weiter ins Detail gehen, lassen Sie uns erst einmal weitere Möglichkeiten sammeln.

## APPENDIX C. DECISION MAKING

VPINO: Was wäre jetzt die *zweite* Lösungsmöglichkeit? Am besten wieder in ein paar kurzen Sätzen.

SIE: die zweite wäre mir hier zuhause einen job zu suchen

VPINO: Und wie würden Sie die zweite Option in einem oder zwei Stichworten abkürzen?

SIE: zuhause

VPINO: So, jetzt haben wir zwei von drei Optionen. Kommen wir zur letzten Möglichkeit.

VPINO: Wie würden Sie diese Alternative beschreiben?

SIE: einfach mal alles auf mich zukommen lassen.

VPINO: Und wie würden Sie diese Option abkürzen?

SIE: abwarten

VPINO: Gut, dann haben wir haben jetzt folgende gesammelt:

Option 1: London

Option 2: Zuhause

Option 3: Abwarten

VPINO: Haben Sie denn schon einen Favoriten? Wenn ja, welcher wäre das?

SIE: wie gesagt, am liebsten würde ich in ausland. also london

VPINO: Also ist Option 1 (London) ihr Favorit?

SIE: ja stimmt

VPINO: Okay.

VPINO: Als nächstes wollen wir mal die Pros und Kontras für die einzelnen Optionen sammeln. Dann beginnen wir am besten gleich mit Ihrem Favoriten, London.

VPINO: Was wäre denn ein erstes Pro für diese Option.

SIE: ich könnte erfahrung sammeln

VPINO: Gibt es weitere?

SIE: hoffentlich gutes gehalt

VPINO: Fällt Ihnen noch ein Pro ein?

SIE: es ist mein favorit :-)

VPINO: Haben Sie vielleicht noch Eins?

SIE: nein das wars

VPINO: Ist denn eines der genannten Pros besonders wichtig?

SIE: ja die erfahrung würde ich sagen

VPINO: Lassen Sie uns über die Kontras von London sprechen. Was spräche denn als erstes gegen diese Option?

SIE: die unsicherheit

VPINO: Fällt ihnen noch ein Kontra ein?

SIE: meine freunde sind alle ganz weit weg

VPINO: Fällt ihnen noch ein Kontra ein?

SIE: ich weiß nicht ob ich überhaupt einen job bekomme

SIE: außerdem is es teuer

VPINO: Gibt es weitere Kontras?

SIE: nein das wars

VPINO: Ist denn eines Ihrer Kontras besonders wichtig? Wenn ja, welches?

SIE: ja das erste

VPINO: Dann fasse ich mal kurz zusammen:

VPINO: Pros:

- ich könnte erfahrung sammeln

- hoffentlich gutes gehalt

- es ist mein favorit :-)

Besonders wichtig: ich könnte erfahrung sammeln

VPINO: Kontras:

- die unsicherheit

- meine freunde sind alle ganz weit weg

- ich weiß nicht ob ich überhaupt einen job bekomme. außerdem is es teuer

Besonders wichtig: Die unsicherheit

VPINO: Auf einer Skala von 1 bis 10, wie hoch schätzen Sie das Risiko für diese Option ein?

SIE: wie risiko

SIE: ?

VPINO: Vielleicht einfach nochmal als Zahl zwischen 1 und 10...

VPINO: Dann fasse ich erneut zusammen:

VPINO: Pros:

- ich könnte erfahrung sammeln

- hoffentlich gutes gehalt

- es ist mein favorit :-)

Besonders wichtig: ich könnte erfahrung sammeln

VPINO: Kontras:

- die unsicherheit

- meine freunde sind alle ganz weit weg

## C.2. Example Decision Support Conversation with VPINO

- ich weiß nicht ob ich überhaupt einen job bekomme. außerdem is es teuer  
Besonders wichtig: die unsicherheit  
VPINO: Auf einer Skala von 1 bis 10, wie hoch schätzen Sie das Risiko für diese Option ein?  
SIE: was meinen sie mit risiko?  
VPINO: Würde es helfen wenn ich einfach mal anders frage?  
SIE: ja  
VPINO: Wie hoch schätzen Sie die Risiken für die Erreichung ihrer persönlichen Ziele ein, die Sie zu Beginn genannt hatten?  
SIE: 9  
VPINO: Also ein eher hohes Risiko.  
SIE: genau  
VPINO: Und wie hoch wären im Vergleich dazu die Chancen bzw. Ihr Gewinn? (wieder auf einer Skala von 1 bis 10)  
SIE: sehr hoch, wenn alles klappt 10  
VPINO: Okay 10  
VPINO: Gut, dann sind wir mit Option "London" erstmal durch.  
VPINO: Reden wir über Option "Zuhause"  
VPINO: Was wäre denn ein erstes Pro für diese Option.  
SIE: alles bleibt beim Alten  
VPINO: Gibt es weitere?  
SIE: ja, ich bleibe bei meinen freunden und der familie  
VPINO: Gibts noch mehr?  
SIE: nein  
VPINO: Ist denn eines der genannten Pros besonders wichtig?  
SIE: ja, das letzte  
VPINO: Folgende hatten Sie genannt:  
Pros:  
- alles bleibt beim Alten  
- ich bleibe bei meinen freunden und der familie  
Welches also meinen Sie?  
  
SIE: das letzte  
SIE: dreunden und familie  
VPINO: Also *ich bleibe bei meinen freunden und der familie.*  
VPINO: Lassen Sie uns über die Kontras dieser Option sprechen. Was spräche denn als erstes gegen diese Option?  
SIE: naja, ich komme persönlich nicht voran  
VPINO: Noch eins vielleicht?  
SIE: ne  
VPINO: Sie haben jetzt nur einen Kontra punkt genannt. Ist dieses Kontra denn besonders wichtig?  
SIE: ja  
VPINO: Dann fasse ich mal kurz zusammen:  
VPINO: Pros:  
- alles bleibt beim Alten  
- ich bleibe bei meinen freunden und der familie  
Besonders wichtig: ich bleibe bei meinen freunden und der familie  
VPINO: Kontras:  
- naja, ich komme persönlich nicht voran  
Das Kontra ist besonders wichtig.  
VPINO: Auf einer Skala von 1 bis 10, wie hoch schätzen Sie das Risiko für diese Option ein?  
SIE: 1  
VPINO: 1 also.  
VPINO: Und wie hoch wären im Vergleich dazu die Chancen bzw. Ihr Gewinn? (wieder auf einer Skala von 1 bis 10)  
SIE: 4  
VPINO: Okay 4, also eher mittel.  
VPINO: Dann sind wir mit Option "Zuhause" erstmal durch.  
VPINO: Reden wir über Option "Abwarten"  
VPINO: Was wäre denn ein erstes Pro für diese Option.  
SIE: eigentlich sehe ich da keine  
VPINO: Haben Sie vielleicht noch Eins?  
SIE: nein  
VPINO: Ist das einzige Pro denn dafür besonders schwerwiegend?  
SIE: nein!  
VPINO: Lassen Sie uns über die Kontras dieser Option sprechen. Was spräche denn als erstes gegen diese Option?  
SIE: das bedeutet ja eigentlich nichts zu tun  
VPINO: Noch eins vielleicht?  
SIE: nein  
VPINO: Sie haben jetzt nur einen Kontra punkt genannt. Ist dieses Kontra denn besonders wichtig?

## APPENDIX C. DECISION MAKING

SIE: nein

VPINO: Dann fasse ich mal kurz zusammen:

VPINO: Pros:

- eigentlich sehe ich da keine

Kontras:

- das bedeutet ja eigentlich nichts zu tun

VPINO: Auf einer Skala von 1 bis 10, wie hoch schätzen Sie das Risiko für diese Option ein?

SIE: 3

VPINO: 3 also.

VPINO: Und wie hoch wären im Vergleich dazu die Chancen bzw. Ihr Gewinn? (wieder auf einer Skala von 1 bis 10)

SIE: 2

VPINO: Okay, dann sind wir mit Option "Abwarten" erstmal durch.

VPINO: Lassen Sie uns doch einmal Ihre Optionen im Vergleich betrachten. Fangen wir an mit London und Zuhause. Die Erste (London) hatten Sie als Favorit genannt. Diese Option klingt ein wenig nach Glücksspiel - große Chance, aber auch ein hohes Risiko. Die andere Option (Zuhause) dagegen bietet eine große Chance bei geringem Risiko. Außerdem haben Ihre Pros überwogen. Das klingt doch eigentlich ziemlich gut.

VPINO: Wie stehen Sie dazu?

SIE: ja stimmt schon, allerdings würde ich nicht sagen dass zuhause bleiben wirklich eine große chance bietet.

VPINO: Und was bedeutet das für Sie als Konsequenz bzgl. Ihrer Entscheidung?

SIE: ja wie gesagt, eigentlich tendiere ich wirklich eher zum in london bewerben

VPINO: Bevor wir nun langsam zum Ende kommen, lassen Sie uns doch noch mal die Option "Abwarten" betrachten.

VPINO: Was sagt denn Ihr Bauchgefühl zu dieser Option?

SIE: : eigentlich ist das ja gar keine option.

VPINO: Diese Option fällt also heraus?

SIE: : eigentlich ja

VPINO: Na gut. Wir kommen nun zum letzten Teil unseres gemeinsamen Dialogs.

VPINO: Sie haben Option1 (London) als Ihren Favoriten genannt. Sind sie dabei geblieben?

SIE: ja

VPINO: Was meinen Sie, ist diese Option für Sie auch real umsetzbar?

SIE: ja das hoffe ich doch

VPINO: Sind Sie immer noch dieser Meinung?

SIE: was?

VPINO: Soll ich die Frage nochmal umformulieren?

SIE: ja

VPINO: Während des Gesprächs hatten Sie behauptet, dass für Sie ein Kompromiss zwischen Ihren Wünschen nicht zur Debatte steht. Nun die Frage, ob Sie noch immer dieser Meinung sind.

SIE: ja, dann sag das doch...

VPINO: Sie sagten zu Beginn, dass Sie sich vor unserem Gespräch noch nicht intensiv mit Ihrer Entscheidung beschäftigt haben. Vielleicht wäre es hilfreich, dass Sie einfach noch mehr Informationen sammeln um einer Entscheidung ein Stückchen näher zu kommen?

SIE: das kann nie schaden

VPINO: Welche Informationen fehlen Ihnen denn noch, um eine Entscheidung treffen zu können?

SIE: ich müsste mich erstmal bewerben

VPINO: Alles klar. Kommen wir zur langsam zum Abschluss.

VPINO: Was sind denn die nächsten Schritte, die Sie nach unserem Gespräch konkret angehen werden?

SIE: firmen in england raussuchen

SIE: und mich bewerben!

VPINO: Okay, ich denke dann sind wir soweit fertig.

VPINO: Ich hoffe wir konnten einer Lösung gemeinsam ein Stückchen näher kommen.

SIE: ja ich denk schon

VPINO: Ich wünsche Ihnen noch einen schönen Tag!

## C.3 Teilnahmebedingungen Nutzerstudie

# Teilnahmebedingungen

## §1 Teilnahmeberechtigung

1. Teilnahmeberechtigt ist jede natürliche Person, die einen festen Wohnsitz in Deutschland hat und das 18. Lebensjahr vollendet hat.
2. Die Verantwortlichen, sowie deren Mitarbeiter und Angehörige sind von der Teilnahme ausgeschlossen.
3. Mit der Teilnahme an der Studie erkennt der Teilnehmer die Teilnahmebedingungen verbindlich an.

## §2 Ablauf

1. Der Amazon-Geschenkgutschein im Wert von 10€ wird Ihnen innerhalb von drei Wochen per E-Mail, an Ihre angegeben E-Mail-Adresse zugeschickt.
2. Hierfür wird Ihr „Abschlusscode“ benötigt, den Sie am Ende Ihrer Teilnahme erhalten.
3. Der Anspruch auf den Amazon-Geschenkgutschein verfällt, wenn Sie nach der Durchführung der Studie nicht innerhalb von 3 Tagen eine E-Mail mit Ihrem Abschlusscode an [vpino@algo.informatik.tu-darmstadt.de](mailto:vpino@algo.informatik.tu-darmstadt.de) senden. Wir empfehlen die Verwendung einer Emailadresse, aus der Ihr Klurname nicht hervorgeht. Weitere Hinweise zum Datenschutz finden Sie hier
4. Ihre Emailadresse wird nur zu diesem Zweck verwendet und nach Abschluss der Studie gelöscht.

## §3 Berechtigung für Erhalt des Amazon-Geschenkgutscheins

1. Um einen Amazon-Geschenkgutschein im Wert von 10€ zu erhalten, müssen Sie alle Fragebögen und das Gespräch gewissenhaft, vollständig und aussagekräftig durchgeführt haben.
2. Eine mehrfache Teilnahme der gleichen Person ist nicht erlaubt und führt zum Ausschluss.
3. Ausgeschlossen sind bzw. werden auch Personen, die sich unerlaubter Hilfsmittel bedienen oder sich anderweitig durch Manipulation Vorteile verschaffen. Gegebenenfalls können in diesen Fällen auch nachträglich Gutscheine aberkannt und zurückgefordert werden.
4. Es wird überprüft, ob diese Bedingungen eingehalten wurden.
5. Die Forscher sind durch die Übermittlung des Gutscheins von sämtlichen Ansprüchen befreit und übernehmen keine Haftung für etwaige Folgeschäden, die aus der Einlösung oder sonstigen Zusammenhängen mit dem Gutschein für Amazon entstehen.

## §4 Amazon-Geschenkgutschein

1. Bzgl. Einlösung und Gültigkeit gelten die Gutscheinbedingungen von Amazon, die auf der Internetseite [www.amazon.de](http://www.amazon.de) eingesehen werden können. Die Einlöse- und Umsetzungsbedingungen sowie die Gültigkeitsdauer der Gutscheine liegen nicht in der Verantwortung des Veranstalters, sondern sind Sache des Gutschein-Gebers (Amazon). Der Gutschein-Gewinner trägt selbst Sorge dafür, sich rechtzeitig und ausführlich über die Einlöse- und Umsetzungsbedingungen sowie die Gültigkeitsdauer beim Gutschein-Geber (Amazon) zu informieren.
2. Amazon.de ist kein Sponsor dieser Studie.
3. Amazon, Amazon.de sind Markenzeichen oder eingetragene Markenzeichen von Amazon EU S.A.R.L. oder ihren verbundenen Unternehmen.
4. Eine Barauszahlung oder ein Umtausch der Gutscheine ist ausgeschlossen.

## APPENDIX C. DECISION MAKING



# Ehrenwörtliche Erklärung<sup>†</sup>

---

Hiermit erkläre ich, die vorgelegte Arbeit mit dem Titel: “Rational Decision Support with a Natural Language Dialogue System” zur Erlangung des akademischen Grades “Dr. rer. nat.” selbstständig und ausschließlich unter Verwendung der angegebenen Hilfsmittel erstellt zu haben. Ich habe bisher noch keinen Promotionsversuch unternommen.

Darmstadt, 07. September 2015

---

Daniel O. Mäurer, Dipl.-Inform.

---

<sup>†</sup>Gemäß § 9 Abs. 1 der Promotionsordnung der TU Darmstadt

## APPENDIX C. DECISION MAKING

# Wissenschaftlicher Werdegang des Verfassers<sup>‡</sup>

---

2003 - 2009	Studium der Informatik Nebenfach: Mathematik Technische Universität Darmstadt
Juli 2009	Abschluß als Dipl.-Inform. Diplomarbeit: "Advanced Speedup-Techniques for Multi Criteria Shortest Path Search in Time-Expanded Graphs" Fachgebiet Algorithmik
2009 - 2015	Wissenschaftlicher Mitarbeiter Fachgebiet Algorithmik Technische Universität Darmstadt

---

<sup>‡</sup>Gemäß § 20 Abs. 3 der Promotionsordnung der TU Darmstadt